NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

UNDERSTANDING FACTORS RELATED TO ATTRITION OF DEPARTMENT OF DEFENSE CIVILIAN EMPLOYEES USING NON-PARAMETRIC SURVIVAL METHODS

by

Brittany Morgan

March 2019

Thesis Advisor:	Lyn R. Whitaker
Second Reader:	Kurt Klingensmith, TRAC Monterey

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE					*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.					
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** March 2019		**3. REPORT TYPE AND DATES COVERED** Master's thesis		
4. TITLE AND SUBTITLE UNDERSTANDING FACTORS RELATED TO ATTRITION OF DEPARTMENT OF DEFENSE CIVILIAN EMPLOYEES USING NON-PARAMETRIC SURVIVAL METHODS				**5. FUNDING NUMBERS** Army Civilian Attrit	
6. AUTHOR(S) Brittany Morgan					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000				**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) OPA, Washington, DC				**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.					
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.				**12b. DISTRIBUTION CODE** A	

13. ABSTRACT (maximum 200 words)

Government success depends on employees with science, technology, engineering, and math (STEM) qualifications to support critical roles within the Department of Defense (DoD), so it is important to understand attrition factors related to the DoD STEM workforce and how these factors might differ from DoD employees in non-STEM occupations. Civilian personnel data from Defense Manpower Data Center (DMDC), linked by anonymous employee identification numbers, was analyzed to study attrition of DoD STEM civilians. To limit the scope of the study, we based the analysis only on a cross-section of civilians employed by Department of the Army (DA) in the first quarter of 2009. Our findings suggest that Virginia STEM employees, especially in the first few years after appointment, have higher attrition rates than the Texas STEM employees. Implications of these results could suggest policy changes such as increases in locality pay for DoD STEM positions in Virginia are needed to incentivize the retention of STEM employees in a competitive market.

14. SUBJECT TERMS attrition, survival analysis			**15. NUMBER OF PAGES** 79
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500 | Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

UNDERSTANDING FACTORS RELATED TO ATTRITION OF DEPARTMENT OF DEFENSE CIVILIAN EMPLOYEES USING NON-PARAMETRIC SURVIVAL METHODS

Brittany Morgan
Lieutenant, United States Navy
BS, U.S. Naval Academy, 2012

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
March 2019

Approved by: Lyn R. Whitaker
Advisor

Kurt Klingensmith
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Government success depends on employees with science, technology, engineering, and math (STEM) qualifications to support critical roles within the Department of Defense (DoD), so it is important to understand attrition factors related to the DoD STEM workforce and how these factors might differ from DoD employees in non-STEM occupations. Civilian personnel data from Defense Manpower Data Center (DMDC), linked by anonymous employee identification numbers, was analyzed to study attrition of DoD STEM civilians. To limit the scope of the study, we based the analysis only on a cross-section of civilians employed by Department of the Army (DA) in the first quarter of 2009. Our findings suggest that Virginia STEM employees, especially in the first few years after appointment, have higher attrition rates than the Texas STEM employees. Implications of these results could suggest policy changes such as increases in locality pay for DoD STEM positions in Virginia are needed to incentivize the retention of STEM employees in a competitive market.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AAG-RFL	Army Analytics Group-Research Facilitation Lab
CAC	Common Access Card
CART	Classification and Regression Tree
DA	Department of the Army
DoD	Department of Defense
DoN	Department of the Navy
DMDC	Defense Manpower Data Center
FCS	Federal Credited Service
FY	fiscal year
GS	General Schedule
IRB	Institutional Review Board
KM	Kaplan Meier
MRA	Minimum Retirement Age
NPS	Naval Postgraduate School
NSPS	National Security Personnel System
OPA	Office of People Analytics
OPM	Office of Personnel Management
PDE	Person-Event Data Environment
PII	personally identifiable information
RAND	Research and Development Corporation
SSN	Social Security Number
STEM	science, technology, engineering, and Math
UIC	Unit Identification Code
ZIP	Zone Improvement Plan

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The Department of Defense (DoD) hires civilian employees in every science, technology, engineering, and math (STEM) field. Government success depends on employees with STEM qualifications to support critical roles within the DoD, so it is important to understand attrition factors related to the STEM DoD workforce and how these might differ for DoD employees with non-STEM occupations (National Academy of Engineering and National Research Council 2012). With civilian personnel data from Defense Manpower Data Center (DMDC), linked by an anonymous employee identification number, and curated and managed by the Army Analytics Group- Research Facilitation Lab (AAG-RFL), we study attrition of DoD STEM civilians. To limit the scope of the study, we base the analysis only on a cross-section of civilians employed by Department of the Army (DA) in the first quarter of 2009. We also show how to "forecast" the proportion of separations for the 2009 cross-section, which is vital to forecasting STEM workforce requirements.

From our survival analysis results, we learn that there is no difference in attrition between males and females in STEM fields, but there is a difference in attrition between males and females in non-STEM fields. This finding is consistent with the study of new hires from Buttrey, Klingensmith, and Whitaker (2018). Despite the lack of differences in attrition of the males and females in STEM positions, geographical differences do have an impact in attrition for STEM employees. We find Virginia STEM employees have higher attrition than the Texas STEM employees, possibly due to the growing STEM job market in Virginia. STEM job availability in Virginia was expected to grow eight percent between 2008 and 2018, according to a 2014 Georgetown University study by Carnevale, Smith, and Melton. The increase in STEM jobs in Virginia is projected to outpace growth in Texas, so there are more STEM job opportunities for the Virginia STEM employees than the Texas STEM employees. These results could lead to policy implications including locality pay adjustments to incentivize retention of these DA STEM employees in Virginia to overcome the competing opportunities the Virginia STEM employees have available to them.

There is a difference in attrition of STEM employees by geographic location, but there does not appear to be much difference in attrition by geographic location (comparing Virginia to Texas) for non-STEM employees. This could be a result of a relatively slow non-STEM job growth rate of 4.5 percent from 2005 to 2015; however, the STEM jobs grew 9.8 percent over the same timeframe (Simmons 2016). Since there appears to be less job mobility for the non-STEM employees, this could explain the lack of difference in attrition trends by geographic location.

The results indicate that age and time until immediately retirement eligible are the most important factors for predicting attrition. This finding is consistent with the life cycle stability hypothesis from the study of Moynihan and Pandey (2007). Their hypothesis suggests that older employees who have been in a position for a long time are negatively correlated with turnover (Moynihan and Pandey 2007). This means that despite the results for Virginia and Texas STEM employees, that overall, age and time until immediately retirement eligible are better predictors of attrition.

References

Buttrey S, Klingensmith K, Whitaker L (2018) Civilian Department of Defense attrition. Technical report TACC-M-TR-18-012, U.S. Army TRADOC Analysis Center, Monterey, CA.

Carnevale A, Smith N, Melton M (2014) STEM state-level analysis. Accessed March 1, 2019, https://cew.georgetown.edu/wp-content/uploads/2014/11/stem-states-complete-update2.pdf.

Moynihan D, Pandey S (2007) The ties that bind: Social networks, person-organization value fit, and turnover intent. *J. of Pub. Admin. Res. and Theory* 18(2): 205–227, doi:10.1093/jopart/mum013.

National Academy of Engineering and National Research Council (2012) *Assuring the U.S. Department of Defense a Strong Science, Technology, Engineering and Mathematics (STEM) workforce.* (The National Academies Press, Washington, DC). https://doi.org/10.17226/13467.

Simmons S (2016) STEM job growth expected to exceed growth in non-STEM jobs over the next decade. Accessed March 5, 2019, https://chmuraecon.com/blog/2016/february/29/stem-growth-expected-to-exceed-growth-in-non-stem-jobs-over-the-next-decade/.

ACKNOWLEDGMENTS

To my husband, Nick: I am grateful for your love, sacrifices, and support over the last two years. I would not have made it through this program without you.

I would like to thank Dr. Lyn Whitaker for her guidance, support, and patience throughout this research. It has been a real privilege working with you.

I would also like to thank Dr. Sam Buttrey, MAJ Kurt Klingensmith, and Dr. Spencer Brien for their guidance and insightful contribution throughout this research.

To United States Army Research and Analysis Center and the Office of People Analytics, thank you for your support through this research.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The Department of Defense (DoD) hires civilian employees in every science, technology, engineering, and math (STEM) field. Government success depends on employees with STEM qualifications to support critical roles within the DoD, so it is important to understand attrition factors related to the DoD STEM workforce and how these might differ from DoD employees with non-STEM occupations (National Academy of Engineering and National Research Council 2012).

The Office of Personnel Management's (OPM) FedScope database provides cross-sectional data and summary statistics for civilian federal employees by year (OPM 2018). Due to the potential Personally Identifiable Information (PII) contained in such data, records are not linked by employee across years. So, although the FedScope datasets give invaluable insight into the federal workforce, they cannot be used to track career paths and attrition. The data sets required to capture civilian DoD career progression, maintained by the Defense Manpower Data Center (DMDC), have not until recently been regularly accessible in secure computing environments that also provide analytical tools (Buttery, Klingensmith, and Whitaker 2018). For this and other reasons, "few federal agencies maintain formal forecasts of STEM workforce requirements" (Butz et al. 2004). Since the agencies are unable to forecast requirements, they are also unable to model attrition. As a substitute for modeling attrition, most studies model turnover intent of federal employees such as studies from Leider, Harper, Shon, Sellers and Castrucci (2016) and Wadsworth, Llorens and Facer (2018). These studies shed light on employees' state view on their likely future action, but do not analyze actual employee behavior.

With civilian personnel data from DMDC, linked by an anonymous employee identification number, and curated and managed by Army Analytics Group-Research Facilitation Lab (AAG-RFL), we study attrition of DoD STEM civilians. To limit the scope of the study, we base the analysis only on a cross-section of civilians employed by the Department of the Army (DA) in the first quarter of 2009. By constructing career paths from the DMDC data, we are able to explore attrition patterns of STEM and non-STEM employees across time using techniques from survival analysis. We also show how to use survival analysis to "forecast" the proportion of separations for the 2009 cross-section, which is vital to forecasting STEM workforce requirements.

A. PRIOR RESEARCH ON FEDERAL GOVERNMENT TURNOVER

There are few studies about federal workforce attrition that are able to measure actual turnover behavior and even fewer concerning civilians working for the DA. In this section, we discuss three studies related to our work. The first study, Copeland (2011), is about characteristics of the federal workforce. The study is relevant because it is based on data from the OPM FedScope database which serves as a comparison to our data. The second study of Nataraj, Hanser, Camm and Yeats (2014) focuses on DA attrition of new hire civilian employees which is relevant because we can draw some parallels of their attrition results to ours. The final study, Asch (2002), looks at federal employees and their estimated survival functions for STEM employees. The data and methods used in this study mirror our data and methods more closely than any other study.

With data from the FedScope database, Copeland (2011), compares data from the years 1998 and 2008 for the entire federal workforce. Despite looking at different characteristics of the federal workforce, the data are not separated by agencies within the federal government. The characteristics Copeland (2011) study include sex, geographic location, blue or white collar, and age distribution. More than 35 percent of the federal employees used in the study are stationed in California, Virginia, Texas, or Maryland. From 1998 to 2008, female representation in the federal workforce only increases from 44.4 percent to 44.7 percent. Despite the slight overall increase of female employees in 2008, the percentage of women in white collar jobs drops from 49.6 percent in 1998 to 48.7 percent in 2008 (with a slight percentage of male increase), and there is only a slight increase of 0.02 percent of female employees (with a slight percentage of male decrease) in blue collar jobs (Copeland 2011). The overall age distribution of the workforce is also older in 2008 than 1998, which means that more of the employees in 2008 are close to retirement.

With data from DMDC and sponsored by Research and Development Corporation (RAND), Nataraj et al. (2014) studied full-time DA civilians who are employed from fiscal year (FY) 2004 through FY 2013. Their goal is to "bring together workforce supply and demand projection models to examine how projected supply might be managed to meet projected authorizations by the end of FY 2017" (Nataraj et al. 2014). Figure 1 shows that the number of full-time Army civilian employees grows until 1987 but then steadily declines and never reaches the maximum number of Army civilian employees attained in the 1980s (Nataraj et al. 2014).

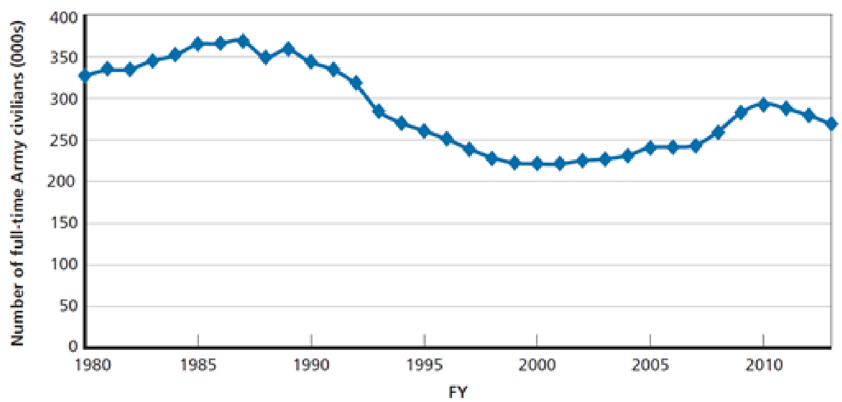

Figure 1. Number of Full-Time Army Civilians. Source: Nataraj et al. (2014).

A comparison of the rates of separations and the rates of new hires for the DA from 2006 through 2013 is in Figure 2 (Nataraj et al. 2014). The separation rates never appear to exceed the new hire rates during this time, and the spike in 2012 is both in new hires and separations.

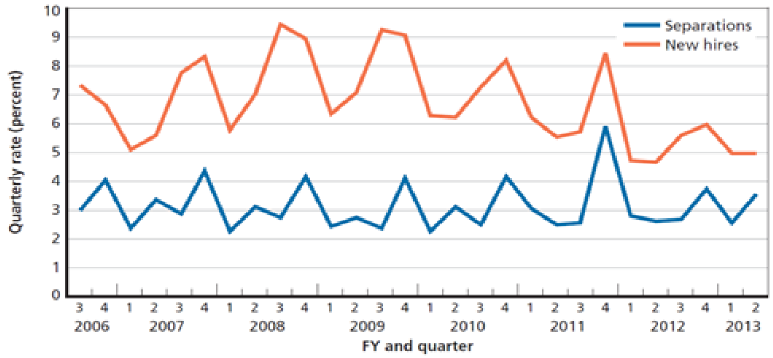

Figure 2. Separation and Hire Rates by Quarter from 2006–2013. Source: Nataraj et al. (2014).

Nataraj et al. (2014) conclude there are competing explanations for the spikes in separations in 2012; among them are base closures, natural attrition, early retirement incentives, and hiring freezes.

A study by Asch (2002) uses survival analysis to understand the retention of General Schedule (GS) employees in the federal workforce. Unlike the Nataraj et al. (2014) study, which quantifies the proportion of separations in each quarter, survival analysis estimates a survival function. Survival functions, often denoted by $S(t)$, are functions of time, t, where Asch (2002) measures t as the number of months since appointment. Generally, for attrition studies, the survival function at time t is the probability that an individual "survives" until time to t (i.e., the probability that an individual does not attrite between times zero and t).

With data from DMDC (2001), two cohorts (FY 1988 and FY 1992) are analyzed for individuals who entered the federal service between FY 1982 and FY 1996 (Asch 2002). This study does not consider males and females separately but keeps them together for analysis. Specifically, looking at the retention of the two cohorts by occupation type, including STEM jobs, the timeframe of observation until separation is watched for five years; the results are shown in Figure 3.

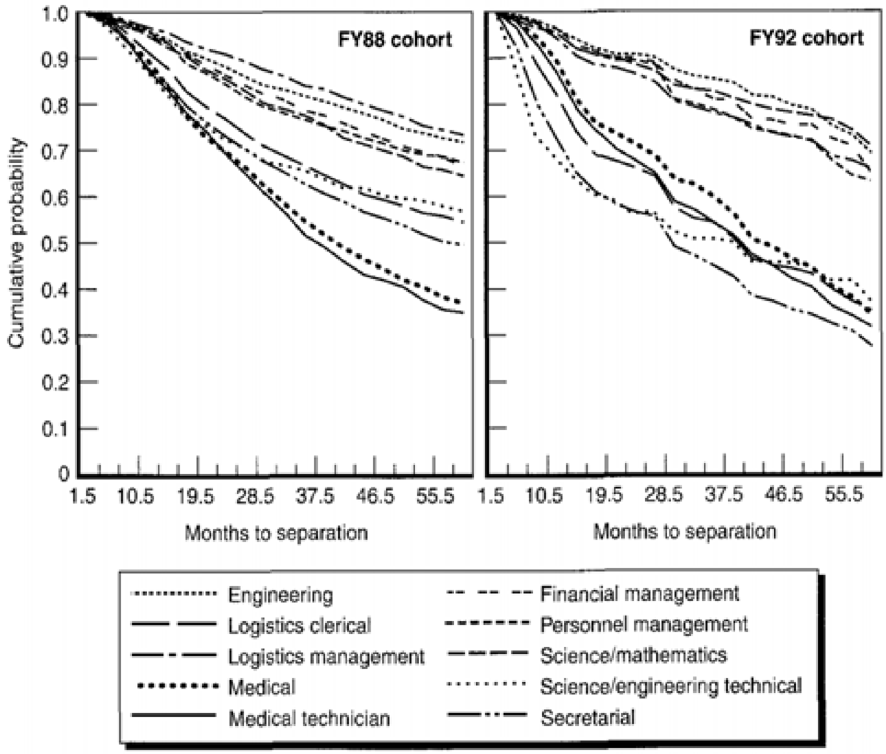

Figure 3. Estimated Survival for FY 1988 and FY 1992 Cohorts. Source Asch (2002).

Although the Asch (2002) study observes new hires over only five years and more closely mirrors the study of Buttrey et al. (2018), there are some similarities to our work. The STEM codes in the Asch (2002) study slightly differ from the STEM codes we use in our thesis (as seen in Appendix A). In our cross-section, the medical Army civilian employees have the worst survival of all of the categories, which mirrors the FY 1988 and FY 1992 cohorts as the medical fields are consistently in the bottom half of the survival curves. Another similarity is with the science/mathematics field in this study: this grouping of employees is the closest grouping to our STEM employees. The low attrition of these science/ mathematics employees in the Asch (2002) study are similar to the STEM employees in our cross-section, who also have low attrition.

B. DATA AND THE LIMITATIONS

We study attrition patterns of STEM civilians employed by DA in the first quarter of 2009 using survival analyses similar to that of Buttrey et al. (2018). To do so, we further focus attention on only full-time, permanently appointed, GS STEM DA civilians and compare them to similar DA civilian employees with non-STEM occupations. Focusing on these employees by excluding wage-grade and temporarily appointed employees makes STEM and non-STEM occupations more nearly comparable. In addition, by excluding temporary appointments, we avoid artificially inflating attrition rates in the first year.

The greatest challenge faced by studies such as ours is the availability of data. Our data, supplied by DMDC via AAG-RFL, includes all DoD civilian master-file quarterly snapshot records on the last day of the quarter from the third quarter of 2005 through the first quarter of 2017 and all available transaction records between 2007–10–30 and the first quarter of 2017. This transaction database is new, and more files are being added to the database. Since the transaction database needs time to mature, we study a cross-section of DA employees who are employed on 2009–03–31 rather than on the earliest date on which both transaction and snapshot records are available. Thus, the study period for the cross-section is from 2009–03–31 to the first quarter of 2017.

A survival analysis of the 2009 cross-section requires an appointment date and the smaller of a separation date and the last date of the study period. The 19.29 percent of employees whose records "disappear" prior to 2009–03–31 without a corresponding transaction file indicating separation are counted as separations. Buttrey et al. (2018) fit survival functions to newly hired employees because appointment date is not included in an employee's quarterly snapshot. To use the entire cross-section, including the 75 percent hired before the third quarter of 2005, we construct appointment dates using a combination of the dates for first observed transaction and snapshot records and the Federal Credited Service (FCS) field. The data and its preparation are described in detail in Chapter II.

C. STUDY APPROACH

Our focus is to study attrition behavior of STEM DA employees and how their behavior might differ from non-STEM DA employees. To accomplish this, we use non-parametric estimates of the survival function to study how attrition varies by employee: age, pay grade, geographic location, if they had prior active duty, and time until eligible for immediate retirement. All variable values are taken from the 2009–03–31 employee snapshot records. In Chapter III, we stratify the cross-section using one to three of these variables at a time and estimating separate survival functions for each strata. From such exploration, we find there is no difference in attrition between males and females in STEM fields, but there is a difference between males and females in non-STEM fields. This finding is consistent with the study of newly appointed employees from Buttrey et al. (2018). We also find the attrition of Texas and Virginia STEM employees is different, with Texas having a higher estimated survival function than Virginia. This difference might be attributable to STEM job availability in Virginia was expected to grow eight percent between 2008 and 2018, according to a 2014 Georgetown University study by Carnevale, Smith, and Melton. The increase in STEM jobs in Virginia is projected to outpace growth in Texas, resulting in more STEM job opportunities for those STEM employees in Virginia.

In Chapter IV, we use a survival tree to both partition the data and to estimate survival functions in each subset of the partition. The results of the survival trees are more difficult to interpret. However, they yield an objective and nonparametric approach to estimating the underlying survival functions. We illustrate in Chapter IV how to use survival tree results to "forecast" the proportion of the 2009 cross-section who will survive an additional t years. Conclusions and suggestions for additional work are given in Chapter V.

THIS PAGE INTENTIONALLY LEFT BLANK

II. DATA

This chapter describes the data acquisition process, the data preparation, and cleaning methods used to construct a cross-section of civilian employees working for DA on 2009–03–31. We also describe the variables that are used to ensure that the cross-section only includes permanently appointed, full-time, GS employees working for the DA on 2009–03–31. The business rules in the data preparation and cleaning portion of Section B are indicators of inconsistencies in the data that analysts must consider when trying to reproduce these results.

A. DATA ACQUISITION

AAG-RFL provides access and support for the data used for this thesis. The data for the thesis resides in the Person-Event Data Environment (PDE). The PDE maintains a high level of data security, so access to the PDE is only available to Common Access Card (CAC) users. In addition, AAG-RFL de-identifies personnel records by replacing social security numbers (SSN) with a randomly generated identification number, which is the variable, "PID_PDE." It further de-identifies certain variables; for example, it masks Zone Improvement Plan (ZIP) codes and scrambles Unit Identification Codes (UICs).

The first step to accessing this data is to request to create a project within the PDE. To have access to the ZIP codes and UICs, AAG-RFL requires a Naval Postgraduate School (NPS) Institutional Review Board (IRB) protocol. This study falls under the NPS IRB approval protocol NPS.2018.0101-IR-EP5A. Once approved, the next step is to look through the data catalog and to identify which tables are necessary for the compilation of the data. Section B of this chapter describes the process after the necessary tables are identified from the data catalog.

B. DATA PREPARATION

The data available in the PDE for this study is stored in 12 Oracle tables. These 12 tables contain data supplied to AAG-RFL by DMDC. They include quarterly snapshot records for all civilian DoD employees (including those who are employed by the DA) taken on the last day of the quarter from the third quarter of 2005 through the first quarter of 2017. They also include records of all transactions (e.g., change of paygrade, separation), for DoD civilians from 2007–10–30 through the first quarter of 2017. Furthermore, so that the effects of prior active duty service and current and prior reserve duty might be captured, these 12 tables also include snapshot and transaction records for all active and reserve duty personnel between 2005 and 2017.

Even though the focus of our study is civilian DA employees, we extract data from all of the tables. The first step is to extract the "PID_PDEs" for a cross-section of all individuals employed by DA on 2009–03–31. For this, we use the DA snapshot file MASTER_CIVAPF_QTR_V3A. Then, we extract all records, transactions, and snapshots for civilian, active duty, and reservists who have those "PID_PDEs." We do this to construct variables that describe the career path of an individual for the available timeframe. This approach will capture, for example, the fact that a civilian worked for the Department of the Navy (DoN) between 2005 and 2007 and then was hired by DA in 2008. This also allows us to construct a complete career path history for civilian employees in the 2009 DA snapshot. For example, the approach will capture the fact that an employee working for DA in 2009 has 20 years of prior active duty service, was hired by DA in 2008, and then, in 2010, the employee left DA civilian employment for a civilian position working for the DoN.

Figure 4 depicts the data cleaning steps to ensure that our analyses are reproducible. We see in the first step in Figure 4 that from the quarterly snapshot files, 278,298 unique "PID_PDEs" are identified. These are the DoD employees who worked for the DA on 2009–03–31. The remaining steps are:

(1) Discard anyone with a first transaction file date more than 93 days or approximately three months before their first snapshot date. We do this because there should not be a disparity between first snapshot date and the first transaction file date.

(2) Discard the few individuals who only have one snapshot record and no transaction records or who have one snapshot record and a transaction record that indicates separation prior to the snapshot record date. These types of records likely indicate individuals with fewer than 90 days of employment with the DA.

(3) Discard anyone who is not in the paygrade of GS2-13 or BAND or has a missing paygrade on 2009–03–31. The paygrade of BAND is defined in Section C.2. The other paygrades removed from the data include but are not limited to non-supervisory pay schedules, executive pay and administratively determined rates.

(4) For the variable "CountryCode," the three levels are United States, Germany and Other. We discard anyone who has a "CountryCode" of Germany but include the levels United States and Other. We do not discard those employees who have a "CountryCode" of Other because we do not know which countries these employees are stationed, so we keep these in the cross-section.

(5) Everyone should have a birthdate; the 14 records with a missing birthdate are discarded. The employees younger than 18 and older than 75 are also discarded.

(6) Seven percent of the cross-section are part-time employees so they are removed and the remaining records are for the full-time employees.

(7) Only the employees who are permanent appointment are included in the cross-section.

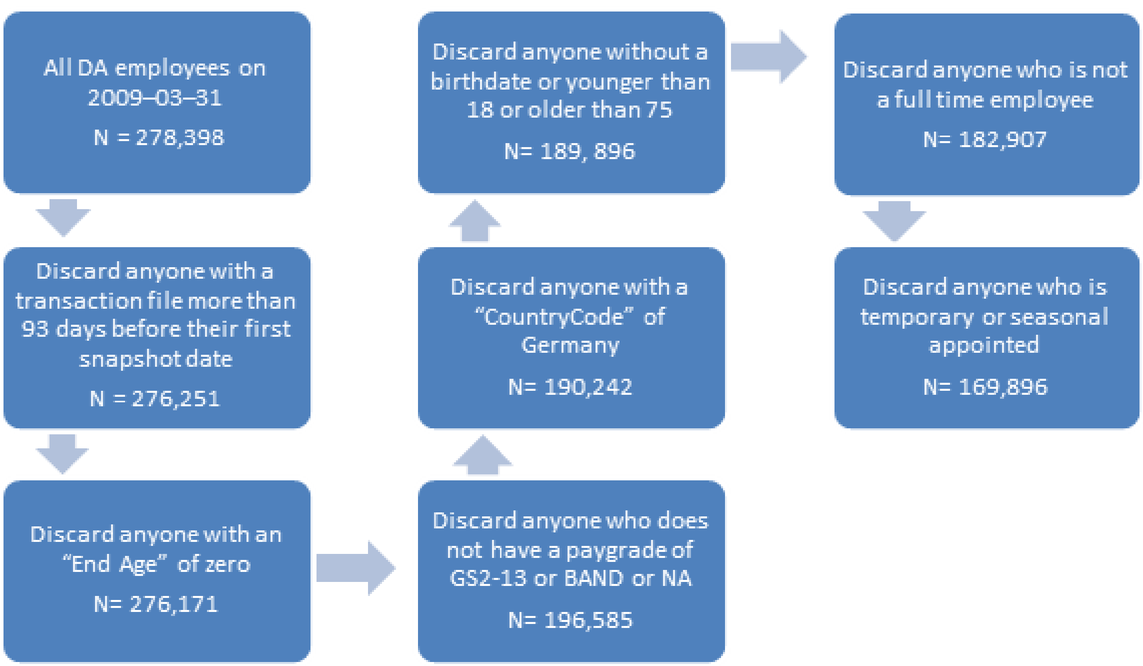

Figure 4. Methodology Flow Chart.

Table 1 shows the distribution of those employees who worked for the DA on 2009–03–31 by their first snapshot date. The employees with first snapshots on 2005–12–31 or later correspond to employees newly appointed to a DA position in that quarter who did not hold a DoD civilian position between 2005–06–01 and the quarter corresponding to their snapshot date.

Table 1. Number of Employees by First Snapshot Date.

≤ 2005–09–30	**2005–12–31**	**2006–03–31**	**2006–06–30**	**2006–09–30**	**2006–12–31**
125,466	3,057	2,703	2,913	2,335	2,015
2007–03–31	**2007–06–30**	**2007–09–30**	**2007–12–31**	**2008–03–31**	**2008–06–30**
1,842	2,855	3,512	3,025	3,283	4,044
2008–09–30	**2008–12–31**	**2009–03–31**			
5,164	3,492	4,190			

Of note, about 75 percent of the cross-section has a first snapshot date on or before 2005–09–30. Most of these people are appointed before the third quarter of 2009. This is because the earliest available snapshot records are for the third quarter 2009, we do not know which or how many of the 125,466 records belong to civilians appointed before 2005–09–30.

C. VARIABLES

Table 2 is a description of the variables in the cross-section. There are five columns in the table which includes: “Name,” “Column Name,” “Modified,” “Type,” and “Description.” The “Name” in the first column is the name we give to the variable. The “Column Name” is the name of the variable in the data. The “Modified” column is a “yes/no” indicator if we modified the variable from Oracle or kept the same values. The “Type” column is the classification of the variable either: categorical, numerical or date. The “Description” column is the description of the variables and is consistent with the ones from Buttrey et al. (2018). These variables are described in more detail in the next three sections.

Table 2. Variables Used in Cross-Section.

Name	Column Name	Modified (Y/N)	Type	Description
Sex	Sex	N	Categorical	Male (M) or Female (F)
Pay Grade	Rank.Code.1	N	Categorical	GS2-GS13 or BAND
Service	SVC.1	N	Categorical	Service Component on 2009–03–31: Army, Navy, Marine Corps, Air Force, DoD
STEM Code	StemCode.1	Y	Categorical	Classification of job in the Department of the Army either: C (Social Sciences), M (Medical), N (Non-STEM) or S (STEM)
Prior Active Duty Status	PriorAD	Y	Categorical	True if prior active duty service or False if no prior active duty service
Years Until Immediate Retirement	YearsIR	Y	Numeric	The number of years until immediate retirement eligible on 2009–03–31
Education	EdFirst	Y	Ordinal	0= less than high school; 1= graduated high school; 2= four-year degree; 3= graduate degree on 2009–03–31
Education Status	EdStat	Y	Ordinal	Track the status of education codes over time 0= data internally consistent; -2= a degree or high school diploma is lost; -20= one is skipped; -22= both of these happened
Birth Date	BirthDate	N	Date	Birth month and year
Age on 2009–03–31	AGE	Y	Numeric	Calculated age from birthdate on 2009–03–31
Work Schedule Type	WK_SCHED_TYP_US_CD	N	Categorical	Full-time position designator on 2009–03–31 with either the values of full-time or part-time
Appointment Code	TYP_OF_APN	N	Numeric	Career designated appointment code
File Date	FILE_DT	Y	Date	First snapshot date for each employee
FCS	FCS_MN_QY	Y	Numeric	Number of months credited as prior service to each employee
FCS Date	FCS Date	Y	Date	Adjusted date to capture total service time in DoD
Start Time	Start Time	Y	Date	Calculated appointment date for each employee
End Age	End Age	Y	Numeric	Age of employment at time of separation or 2017–03–31, whichever occurs first
Event	Event	Y	Categorical	Whether the employee separated prior to 2017–09–30 or not

D. CATEGORICAL VARIABLES

In this section, we define variables that are categorical in our cross-section.

1. SEX

Of the DoD employees employed by DA on 2009–03–31, 44.78 percent are females, and 55.22 percent are males. The percentage of females in our cross-section are comparable to the percentage of women in the total federal workforce in 2012, which is 44.2 percent (Jeffrey 2014).

2. PAY GRADE ON 2009–03-31

"Within the Federal service, the GS classification and pay system cover about 1.5 million positions. The GS has fifteen grades from GS1 (lowest) to GS15 (highest)" (OPM 2018). The cross section does not have any GS employees with either a GS1 grade or GS14–15 grade. Pay grade is usually related to education level; for example, those applicants with a high school diploma qualify for GS2 jobs, those applicants with a bachelor's degree qualify for GS5 jobs, and those applicants with a master's degree qualify for a GS9 job (OPM 2018). Each GS grade has ten step rates at which an employee can receive approximately a 3 percent pay raise. The data do not include information about step rates for any federal employee. Having the steps within a GS level would be beneficial to track promotions. The particular steps are not available through the PDE. However, this data might be available in the original DMDC files.

During 2006–2009, there is a short shift from the GS pay scale to the National Security Personnel System (NSPS), also known as pay banding. Though Congress repealed pay banding in 2009, the full repeal of NSPS was not until 2010 for the DoD. As the cross-section includes Federal employees who have a snapshot in 2009, there are some employees on the GS pay scale and some on the pay band scale.

Table 3 is the percentage of the employees in each pay grade. Of note, the smallest percentages are GS2 and GS3 but also GS10 and GS8, which are pay grades that do not have any STEM employees in this cross-section.

Table 3. Percentage of Employees in Each Pay Grade.

Pay Grade	Percentage
GS2	0.046
GS3	0.22
GS4	3.45
GS5	6.61
GS6	5.56
GS7	9.40
GS8	1.64
GS9	8.80
GS10	1.06
GS11	10.19
GS12	9.85
GS13	5.56
BAND	37.62

3. WORK SCHEDULE TYPE

Different work schedule types include full-time, part-time, and seasonal. As the goal is to study employees with similar characteristics, the employees who are not full-time are removed from the data.

4. STEM CODE

The "STEMCode.1" variable is an employee's job classification on 2009–03–31. The four categories are STEM, non-STEM, social sciences, and medical. The percentage of each "STEMCode.1" is in Table 4 and the codes associated with the STEM, medical, and social sciences fields are in the Appendix A.

Table 4. Percentage of Employees in Each STEM Code.

STEM Code	Percentage
STEM (S)	17.33
Non-STEM (N)	77.11
Social Sciences (C)	0.83
Medical (M)	4.73

5. PRIOR ACTIVE DUTY SERVICE

From the cross-section, there are 14 percent with prior active duty service. Table 5 is the number of prior active duty service members and no prior active duty service by STEM and non-STEM. This variable is included because of the influence from the Veterans' Preference Act of 1944. This act is to ensure veterans' preference when competing for either temporary or permanent positions in the federal government (Vocational Rehabilitation and Employment Manual 2014).

Table 5. Number of Employees by Status of Prior Active Duty Status in STEM or Non-STEM Fields.

	Prior Active Duty Service	No Prior Active Duty Service
STEM	1,699	27,747
Non-STEM	21,021	109,994

E. NUMERIC VARIABLES

In this section, we define variables that are numeric in our cross-section.

1. EDUCATION ON 2009–03–31

As each GS position has education requirements, an employees' education should be adequately tracked throughout their career. Fortunately for this cross section, 97.61 percent of the education records are sequential; for example, an employee who has a master's degree is also showing both a high school diploma and a college degree. Approximately 1.56 percent of the cross section lost track of either a high school diploma or a college degree. About 0.69 percent of the cross section has an education record skipped in their file. Approximately 0.13 percent of the cross-section has both a record skipped and a degree or diploma lost.

The education level of the cross-section reflects in Table 6. The data does not indicate if an employee has a two-year degree, nor does it reflect if a federal employee has a doctorate.

Table 6. Percentage of Employees by Highest Recorded Education Level on 2009–03–31.

Education Level	Percentage
Less than High School	0.47
Graduated High School	59.27
Four-Year degree	26.92
Graduate Degree	13.33

2. AGE ON 2009–03–31, AGE GROUP

The calculation of age for each employee is the difference between 2009–03–31 and their birthdate. The purpose of this calculation is to determine the ages (to the nearest month) of the employees on 2009–03–31, from the given birth date. The age distribution in Figure 5 appears to be fairly uniform between the ages of 27 and 36, with a dramatic increase, followed by another uniform section between the ages of 46 and 55. Almost 75 percent of full-time, permanently appointed, DA civilians in the 2009–03–31 cross-section are older than 40.

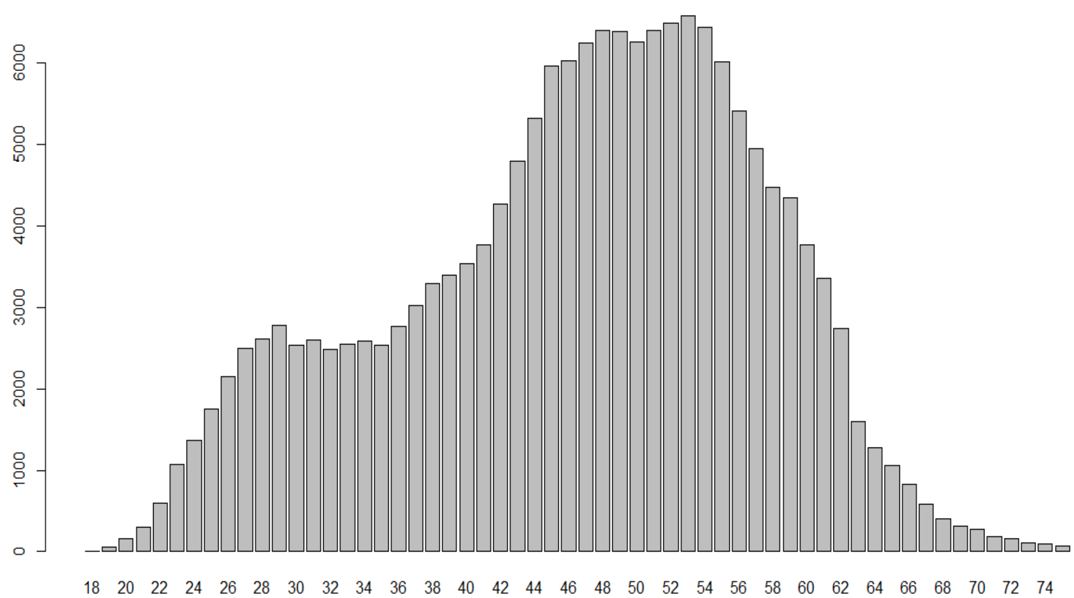

Figure 5. Frequency Distribution of Ages on 2009–03–31.

For analysis of the ages of the cross-section, we also group ages into classes that have about 10 percent of employees in each group. The percentage of employees in each group are in Table 7.

Table 7. Percentage of Employees by Age Group.

[18,30]	(30,37]	(37,42]	(42,45]	(45,48]	(48,50]	(50,53]	(53,56]	(56,59]	(59,75]
10.52	10.92	10.74	9.45	10.99	7.44	11.45	10.50	8.10	9.90

3. TIME TO IMMEDIATE RETIREMENT ELIGIBILITY ON 2009–03–31

OPM defines three different immediate retirement conditions, and if any of these conditions are met an employee is eligible to draw an annuity immediately upon retirement. The first immediate retirement condition is that an employee must have a minimum age of 62 years old and five years of federal service. The second immediate retirement condition is a minimum age of 60 years old with 20 years of federal service. The third immediate retirement condition is a Minimum Retirement Age (MRA) with 30 years of federal service where MRA is between 55 and 57, depending on year of birth. We approximate MRA to be 55 if the employee is born before 1952, or 56 if the employee is born between 1953 and 1969, or the MRA is 57 if the employee is born after 1970 (OPM 2019).

4. APPOINTMENT TYPE

There are five classes of appointment type. The type of appointment varies from permanent or career to limited or non-career designated. The appointment types corresponding to career designated employment are listed in Table 8. All other appointment types not consistent with career designated employment, are removed from the data. The remaining types and the frequency of each are in Table 8.

Table 8. Type of Appointment for Each Employee.

Value	Description	Frequency
1A	COMPETITIVE—CAREER	127,287
1B	CAREER EXECUTIVE ASSIGNMENT—CAREER	1
1C	EXCEPTED—CAREER	13,168
2A	COMPETITIVE—CAREER CONDITIONAL	19,209
2B	CAREER EXECUTIVE ASSIGNMENT—CONDITIONAL	4
2C	EXCEPTED—CAREER	5,091
2D	CANAL ZONE—CONDITIONAL	1
2F	VETERANS READJUSTMENT APPOINTMENT	5,135

F. TIME AND AGE VARIABLES

In this section, we define variables that are dates or are computed based on dates.

1. BIRTHDATE

Birthdate is represented in the PDE only as the year and month of birth and is used to compute an individual's age on 2009–03–31. The 14 individuals without a birthdate are removed from the data.

2. FILE DATE

The first file date for employees who are hired after 2005–09–30 is the date of their first snapshot. For the employees who are hired on or before 2005–09–30, the first file date is reported as of 2005–09–30 even though employees are hired before that time.

3. FEDERAL CREDITED SERVICE

There are employees who have no appointment date. For these employees with no appointment date, we use a surrogate date which is constructed by using their FCS date. The FCS date describes the amount of prior federal service an employee has had, adjusted to correct for discontinuities in service and/or credit from; for example, active-duty military service. Prior federal service includes service to DA, other DoD components and service to any other federal agency. The “FCS_MN_QY” variable gives the number of months credited as prior service to each employee. That variable is used in the computation of “Start Time” as described in the next section.

4. START TIME, START AGE

“Start Time” is a computed date meant to approximate the appointment date. For those employees with a first file date after 2005–09–30, the “Start Time” is the first file date, even if there was a non-zero “FCS_MN_QY” or an earlier transaction, with the exception of the business rule from Chapter II, that if the first transaction is more than 93 days earlier than the first file date, the employee is dropped from the cross-section. For those employees whose first file date is 2005–09–30 or earlier, the “Start Time” is computed as the first file date minus the first value of “FCS_MN_QY.” Since that FCS quantity is in months, it is converted to days by multiplying the length of the average month, which is 30.44 days. Notice that this approach requires us to treat the truncated employees, who start on or before 2009–09–31, differently than we do the employees who start after.

For the variable “Start Age,” age refers to the number of years employed at “Start Time.” For instance, for those employees who start on 2009–09–31, their “Start Age” is zero because they start on the date we start watching. Unlike more recently hired employees, employees whose “Start Time” is computed using FCS date will have “Start Ages” that include non-DA and non-DoD federal service.

5. END TIME, END AGE

“End Age” is the time (in years) that represents either the time served as a civilian DA employee or those who separate prior to 2017–09–30. For those who are still working for DA on 2017–09–30, “End Age” represents the time served as a DA civilian employee up to 2017–09–30. In fact, we cannot compute this value exactly for all employees because everyone works for DA on 2009–03–31, but we credit them employment for other DoD components when computing “End Age.”

6. EVENT

An event is True when someone separates from the DA. If the event is True, their last transaction is either "separated" or "disappeared." The last transaction detail of "separated" is consistent with the definition that OPM provides in their reference, but "disappeared" is not defined in the OPM reference. In Buttrey et al. (2018), "disappeared" is defined as anyone whose snapshots disappear, and for these employees, there are no transaction records indicating that the disappearance is caused by a separation from federal service. Those employees with an event of true and "separated," while we are watching them, are 49.73 percent of the cross-section, and those who do not separate while we are watching are 50.27 percent of the cross-section.

THIS PAGE INTENTIONALLY LEFT BLANK

III. EXPLORATORY SURVIVAL ANALYSIS

We use survival analysis to estimate the survival function *S(t)* or the probability that a civilian separates after time *t*, as a function of *t*. The survival time of a DA civilian employee is defined by their tenure serving as a civilian employee for the DA from appointment date up to separation or 2017–09–30, whichever occurs first. Attrition or the "death" of a DA civilian is defined as when a civilian employee separates from federal service prior to 2017–09–30. We take a non-parametric approach to estimating survival functions rather than a parametric one because it is unlikely that survival functions from different groups of civilian employees belong to a single parametric family.

A. CROSS-SECTION AND SURVIVAL ANALYSIS

The data for analysis are a cross-section of full-time, permanently appointed, GS employees who are employed by the DA on 2009–03–31. For simplicity, in what follows, we call this data set the cross-section. For this cross-section, we have both transaction records and snapshot records. Our snapshot records go back to 2005, but our transaction records start in 2007. Survival times for the individuals in our cross-section may be "left-truncated" or "right-censored." Figure 6 depicts hypothetical appointment dates and separation dates of seven individuals where the blue x's represent appointment and separation dates that occur between 2005 and 2017. The center vertical line represents 2009–03–31, which is the date everyone in the cross-section works for the DA. Individuals A, B, C, and G are included in the cross-section because they are employed by DA on 2009–03–31. Individuals D, E, and F are not included in the cross-section, even though separation dates are available for all three individuals and appointment dates are available for individuals D and F. Individuals C and G are left-truncated because they are hired before 2009. In addition, we do not know the exact appointment date for individual G, so we approximate the "Start Age," as described in Chapter II, using their FCS date. Individual B is right-censored because we only have snapshot files until 2017, and therefore we do not know about their records after 2017. Both right-censoring and left-truncation must be accounted for when estimating survival functions because the consequences of not doing so biases the estimates.

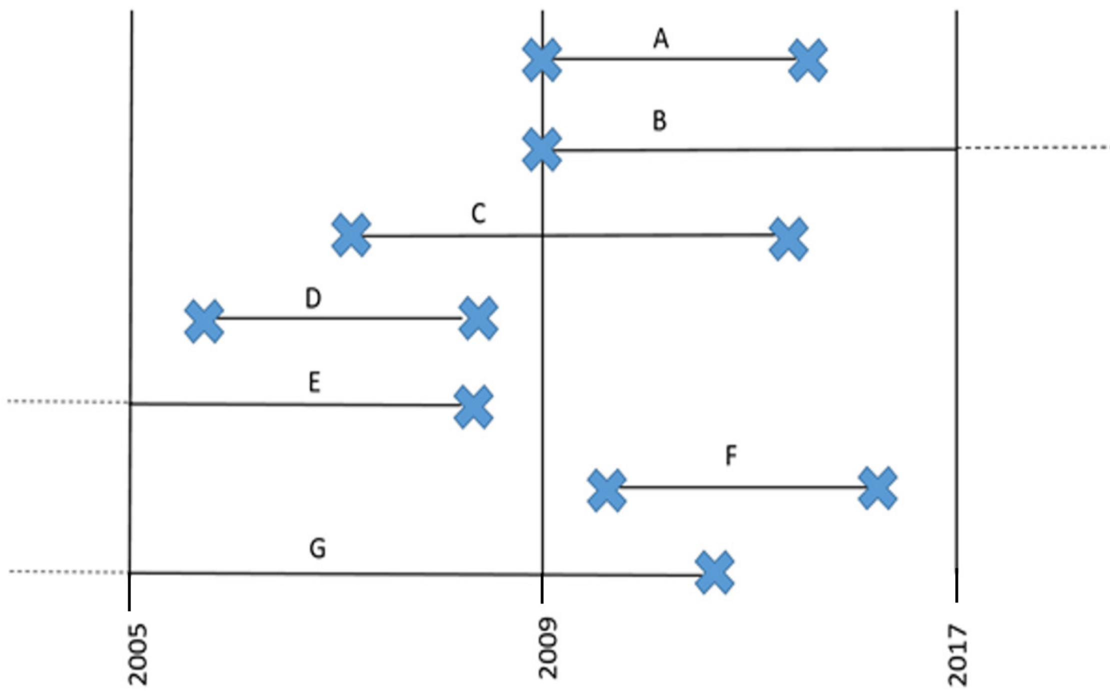

Figure 6. Example of Survival with Left Truncation and Right Censoring.

The non-parametric method of Kaplan and Meier (1958) is used for the analyses in this chapter. The Kaplan Meier (KM) estimator estimates *S(t)* based on data that may be left-truncated and/or right-censored. Specifically, the KM estimator produces a curve where the x-axis is the time (in this case time is in years since appointment), and the y-axis gives estimates of *S(t)* that an individual survives *t* years after appointment. For more details about survival analysis in general and the KM estimator, reference Kleinbaum and Klein (2005).

B. SURVIVAL ANALYSIS RESULTS

Only 49.73 percent of those employees who work for DA on 2009–03–31 separate before 2017. The remaining 50.27 percent are still employed by DA in the first quarter of

2017. Figure 7 displays the estimated survival function with 95 percent confidence intervals for the entire cross-section. The narrowness of the intervals indicates that the cross-section has a large number of observations. It is not a legitimate 95 percent confidence interval because it does not account for the extra uncertainty due to approximating appointment dates for employees appointed before 2005–09–30.

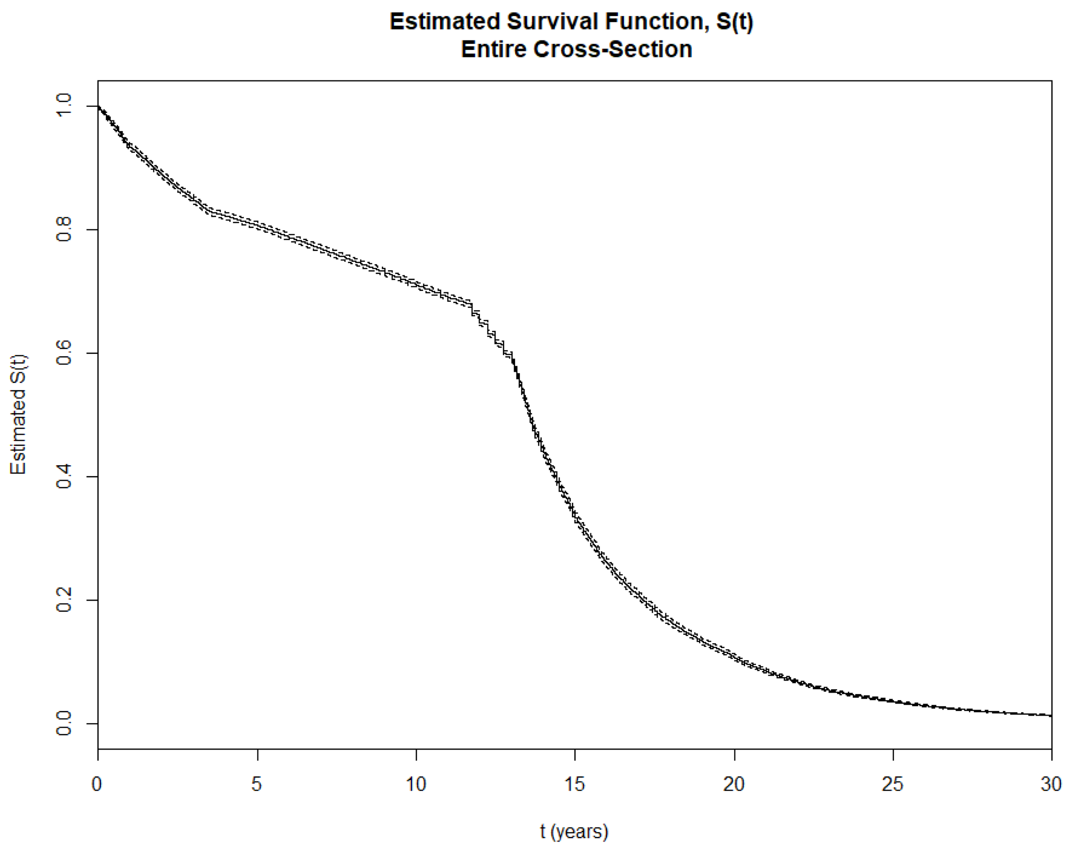

Figure 7. Estimated Survival Function with Confidence Interval for Cross-Section.

In Figure 7, for t less than about 11 years, the estimated probability that a full-time, permanently appointed, GS employee is still employed appears to be piecewise linear over three time intervals: 0–1, 1–4, 4–11. The steepness of these pieces indicates a drop in attrition rate at one year and then at four years after appointment. Attrition rate is directly related to the hazard rate of a survival function. For t where $S(t)$ is continuous, it is defined as,

$$h(t) = -\frac{\frac{d}{dt}S(t)}{S(t)},$$

and is interpreted as the instantaneous likelihood of failure given survival to time *t*. The decrease in attrition rate at the one-year mark and four-year mark in Figure 7 seem plausible. We expect higher attrition in the first year for employees in a probationary status. It is also reasonable to expect higher attrition in the first four or five years after appointment than in the next four or five years.

The stair steps of the estimated *S(t)* in Figure 7 for *t* between 11 and 13 years are unexpected. After 13 years the estimated survival function smooths out but looks quite different from the estimated *S(t)* for *t* less than 11 years. Because we only have snapshot records for about a 13-year window, the estimated *S(t)* for times greater than 13 years depend entirely on our computation of "Start Age." As discussed in Chapter II, "Start Age," is an approximation of the number of years employed by DA on 2009–03–31 based on FCS dates for appointments prior to 2005. It may very well be the case that the abrupt change in Figure 7 at time 11 is due, in part, to how "Start Age" is approximated. Therefore, as civilian employment data accrues the beginning of the stair-steps in Figure 7 should shift to the right.

The shape of the estimated survival function for times greater than 13 years also looks plausible. As employees age, and get closer to retirement, we expect the attrition rate to start increasing. For exploratory purposes, it is valuable to compare the shapes of the estimated *S(t)* for different subsets of the cross-section, for *t* less than 11 years and separately for *t* greater than 13 years. However, because appointment date is approximated, care must be taken in ascribing too much meaning to the values of the estimated *S(t)* for *t* greater than 13 years.

In the remainder of this section, we explore differences in survival functions for different subsets of the cross-section. These subsets are based on features of the civilian employees as recorded in their 2009–03–31 snapshots. Some variables, such as "Sex," "Prior Active Duty Service," and "STEM code" do not tend to change over time. Other variables such as "Age" and "Time Until Immediate Retirement Eligibility" change in predictable ways, and still other variables such as "Pay Grade," and to a lesser extent, "Education Level" vary in time. For exploration in Chapter III, we do not account for changes in these variables over the course of an employee's career.

1. STEM AND GENDER RESULTS

For the STEM employees in the cross-section, there are 22.1 percent females and 77.9 percent males. The estimated survival functions for the STEM, non-STEM and medical employees are in Figure 8. Social sciences are not displayed in Figure 8 because they make up less than one percent of the cross-section, so their results could be misleading for analysis.

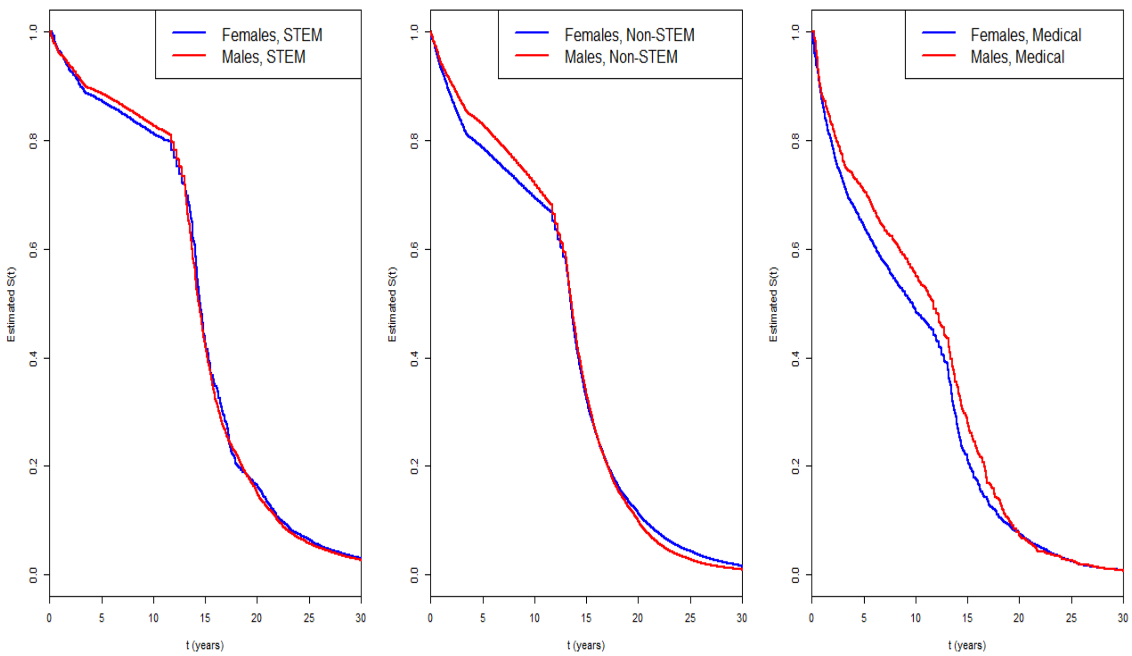

Figure 8. Estimated Survival Functions by Sex for STEM, Non-STEM and Medical Employees.

The estimated survival functions for gender vary by STEM, non-STEM and medical employees. For all three of the different STEM codes, there is a drastic drop after year 11, which is consistent with the estimated survival function for the entire cross-section in Figure 7. As seen in Buttrey et al. (2018) for new hires, there does not appear to be a difference between males and females for STEM. For non-STEM, there appears to be a difference in the first 11 years but then not again. The biggest difference is for the males and females in the medical fields which is interesting because there are 6,221 females in medical and only 1,807 males in medical. The males in this cross-section are also more highly educated than the females, which could be another reason why the female attrition is higher than male attrition in the medical field.

2. STEM AND GEOGRAPHIC LOCATION RESULTS

Consistent with the study from Copeland (2011), the two states with the greatest numbers of DA employees are Virginia and Texas. The number of non-STEM employees in Texas and Virginia is 12,221 and 13,230, respectively, and the number of STEM employees in Texas and Virginia is 1,704 and 2,580, respectively. The median ages of the STEM employees in Texas and Virginia are similar at 49 years old and 50 years old, respectively.

Table 9 gives the number of males and females in each field by geographic location. There are about the same number of males as females in non-STEM in Texas and Virginia. However, the number of males in STEM occupations in both Texas and Virginia is more than double the number of females. Furthermore, the distribution of STEM codes differs between Texas and Virginia. Virginia has a higher proportion of STEM jobs for both males and females.

Table 9. Number of Males and Females by Job Code and Duty Location and Proportion in STEM codes by Sex and Location in Parentheses.

Texas				
	Medical	**Non-STEM**	**STEM**	**Total**
Female	1064 (0.15)	5880 (0.80)	391 (0.05)	7335
Male	353 (0.04)	6341 (0.79)	1313 (0.16)	8007
Virginia				
	Medical	**Non-STEM**	**STEM**	**Total**
Female	381 (0.05)	6906 (0.86)	782 (0.10)	8069
Male	79 (0.01)	6324 (0.77)	1798 (0.22)	8201

From Figure 9, there is not much of a difference in the estimated *S(t)* of Virginia and Texas for non-STEM and medical employees. The interesting difference is in the estimated *S(t)* of the two states for STEM employees. The estimated *S(t)* of the Virginia employees remains higher than the Texas STEM employees for the entire 30 years. The Virginia STEM attrition rate is particularly high in the first three years as can be seen in Figure 9.

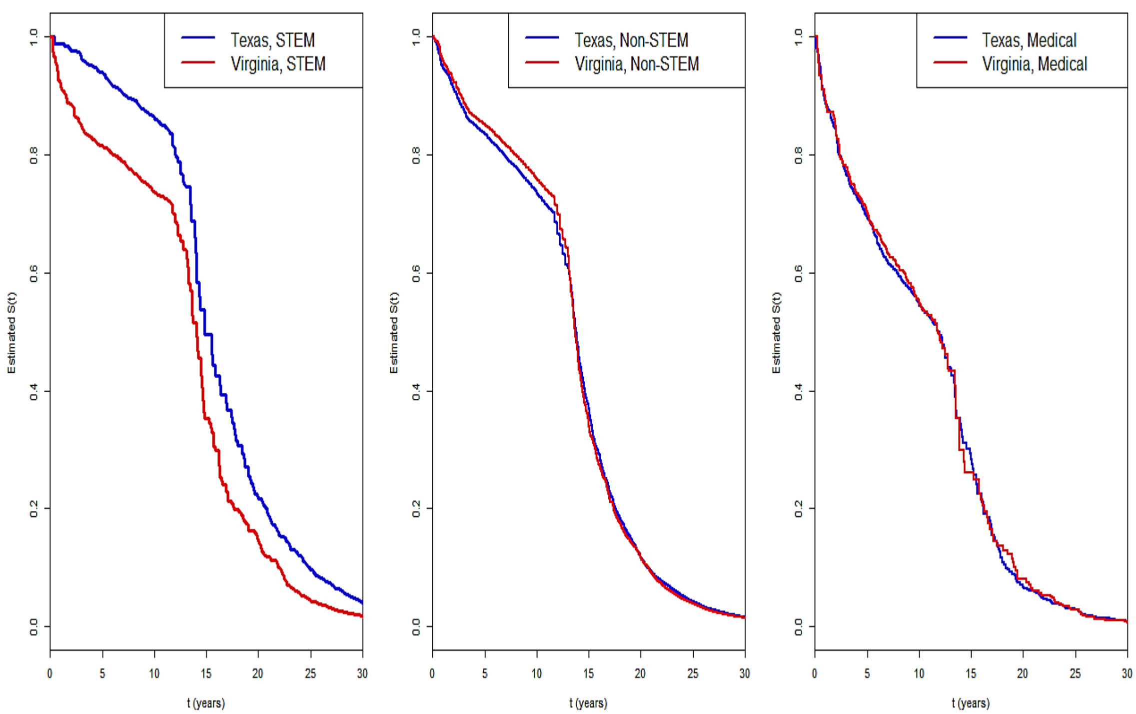

Figure 9. Estimated Survival Functions by Duty Location for STEM, Non-STEM and Medical Employees.

An explanation for higher attrition of STEM employees in Virginia than in Texas could be due to the location of the STEM jobs in each state. As seen in Figure 10, the density of the STEM jobs in Virginia is higher than the density of the Texas STEM jobs. Since the STEM jobs in Virginia are closer, there is greater mobility for these employees than in Texas due to lower costs or hardships associated with seeking and transitioning to other STEM job opportunities within the state.

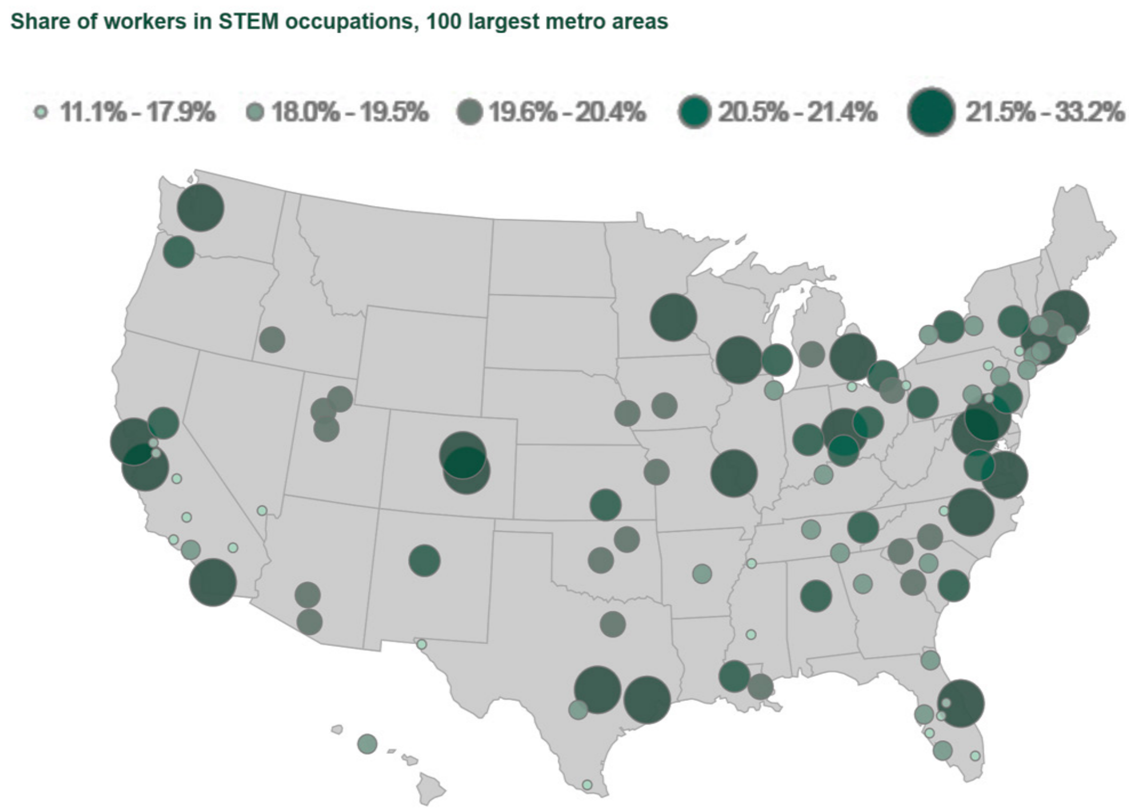

Figure 10. Share of Workers in STEM Occupation by 100 Largest Metro Areas. Source: Rothwell (2013).

In Figure 11, we plot the estimated conditional survival functions among Texas and Virginia STEM employees who have survived the first three years of DA appointment. Because all employees in this group have survived the first three years, the conditional survival function is one for t less than three and $S(t)/S(3)$ for t at least three where $S(t)$ is the unconditional survival function. For t less than about 13 years, the estimated conditional survival functions in Figure 11 are almost the same. For t between about 13 years and 18 years, the estimated $S(t)$ for the Virginia STEM employees drops faster than for Texas STEM employees. The differences we see in Figure 9 are mostly in the first three years. Even so, the differences in the first three years could be due to a projected increase in STEM jobs in Virginia by eight percent over the time of 2008 to 2018 (Carnevale et al. 2014).

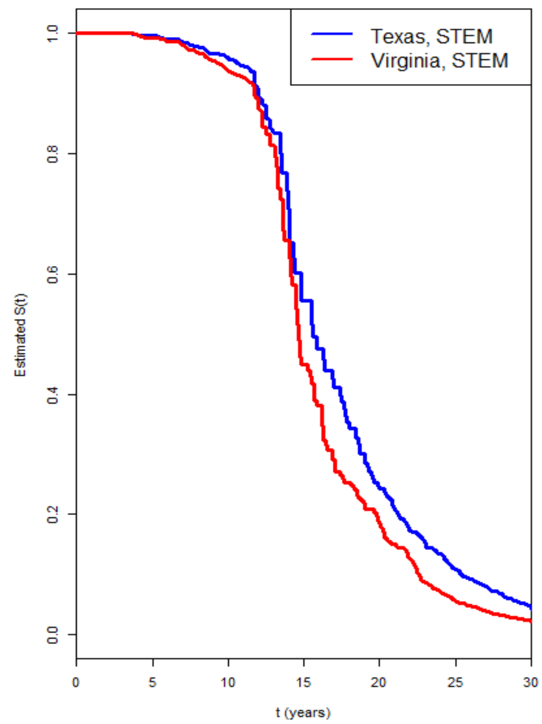

Figure 11. Estimated Survival Functions by Duty Location for STEM Employees on 2011–03–31.

The estimated survival functions for Figure 12 are for STEM employees by the variables "Duty Location" and "Pay Grade" on 2009–03–31. STEM employees in Figure 12 only have a GS pay grade of GS09, GS11, GS12, and GS13. As seen in Appendix B, the largest proportion of these employees are in the BAND pay grade. As mentioned in Chapter II, we are not sure of the exact pay grades of the employees in the BAND pay grade but because they account for more than 30 percent of the STEM employees in Virginia and Texas, the BAND estimated survival functions are similar to the STEM estimated survival function in Figure 9, so we display them separately in Figure 12.

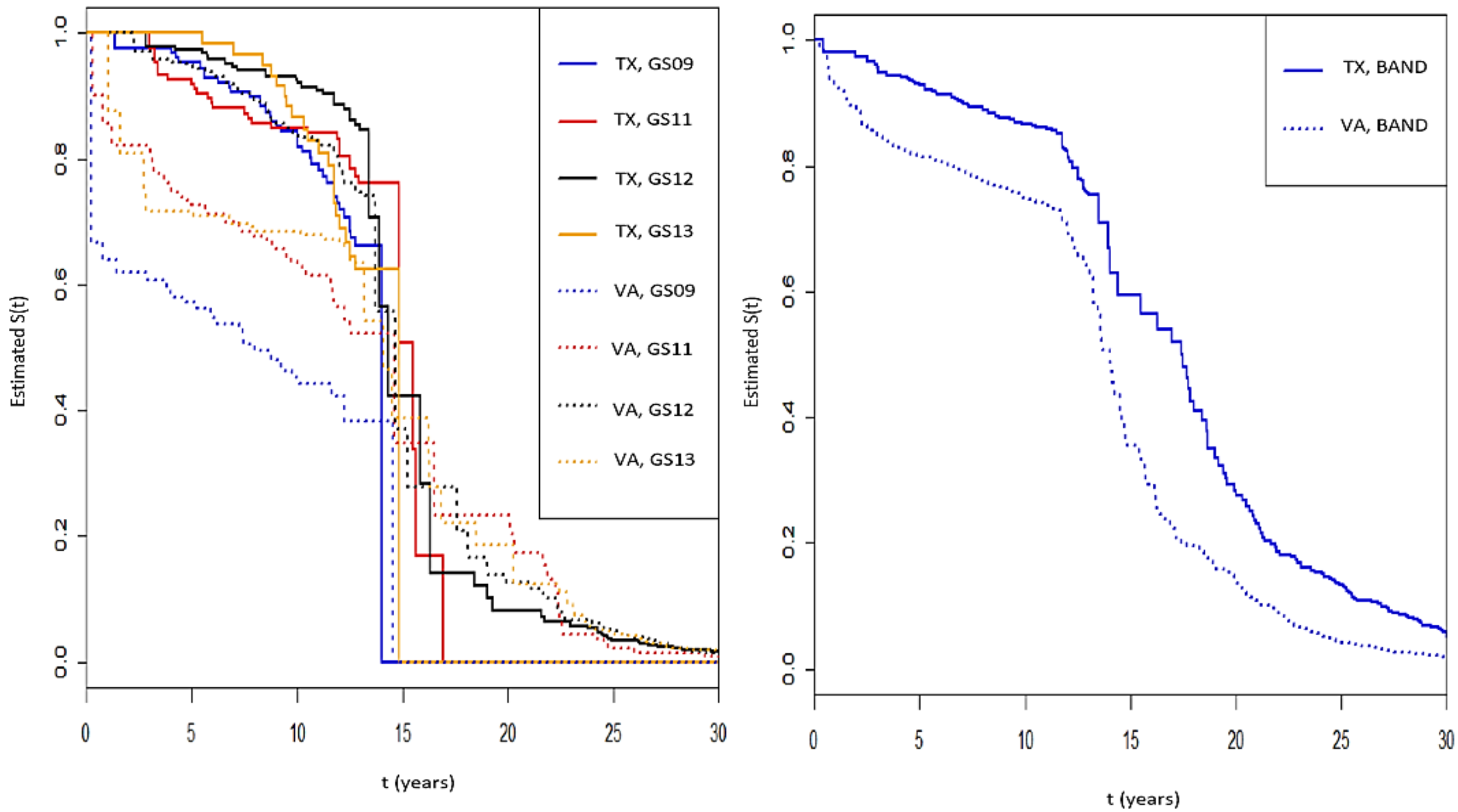

Figure 12. Estimated Survival Functions by Duty Location and Pay Grade for STEM Employees on 2009–03–31.

From Figure 12, we can see that the STEM employees in Virginia have lower estimated survival functions in every pay grade than the STEM employees in Texas. The difference could be that in Texas more than half of the employees have more than 10 years until immediate retirement eligible, more of the employees were new hires, than in Virginia from 2006 to 2009.

3. STEM AND EDUCATION RESULTS

Although we have records of employees with less than high school education, they account for fewer than one percent of the cross-section, so those employees are not considered in this section.

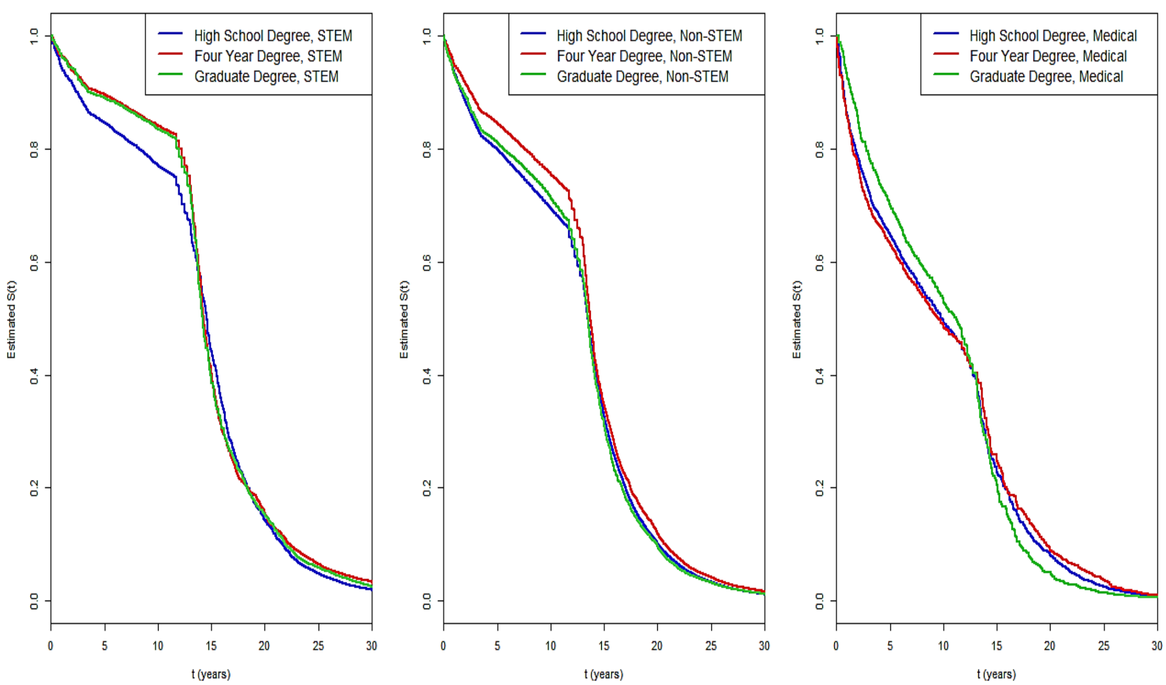

Figure 13. Estimated Survival Functions by Education Level and STEM, Non-STEM and Medical Employees.

From Figure 13, we see that the STEM employees survive with a higher probability than the non-STEM employees. The STEM employees with either a four-year degree or a graduate degree also survive longer than the STEM employees with a high school degree.

The non-STEM employees have similar estimated survival probabilities across all educational levels, but it appears that the non-STEM employees with a four-year degree have a higher estimated survival function than the non-STEM employees with either a graduate or a high school diploma. The higher attrition of the non-STEM employees with a graduate degree could be due to more job opportunities available to them.

For medical employees, Figure 13 displays the employees with graduate degrees have a higher probability of surviving than those with lower degrees until about the 12-year mark. Unlike the STEM and non-STEM employees who have slowly decreasing estimated survival functions, the medical estimated survival function steadily decreases over the 30 years.

4. STEM AND IMMEDIATE RETIREMENT ELIGIBILITY RESULTS

The number of years until immediate retirement eligibility is almost perfectly (negatively) correlated with "Age" for employees under the age of 50. To better understand the effects of nearing retirement and to try to decouple those effects from age, the years are grouped in Table 10. The 10 percent of employees who have a negative time until retirement have reached immediate retirement eligibility. The largest proportion of the employees in the cross-section have at least 10 to 20 years (from 2009–03–31) until they are eligible for retirement.

Table 10. Percentage of Employees of Years until Retirement by STEM

	[-19,0]	**(0,1]**	**(1,2]**	**(2,3]**	**(3,4]**	**(4,5]**	**(5,10]**	**(10,38]**
STEM	1.8	0.39	0.66	0.46	0.45	0.56	2.2	7.9
Non-STEM	9.6	2.4	3.9	2.9	2.8	3.1	13.4	47.4

Overall, from Figure 14, the STEM employees survive better than the non-STEM employees. The non-STEM employees who are already retirement eligible and furthest from retirement eligible have the worst survival. There is still a drastic drop around the 10-year mark which is consistent with Figure 7 of the entire cross-section.

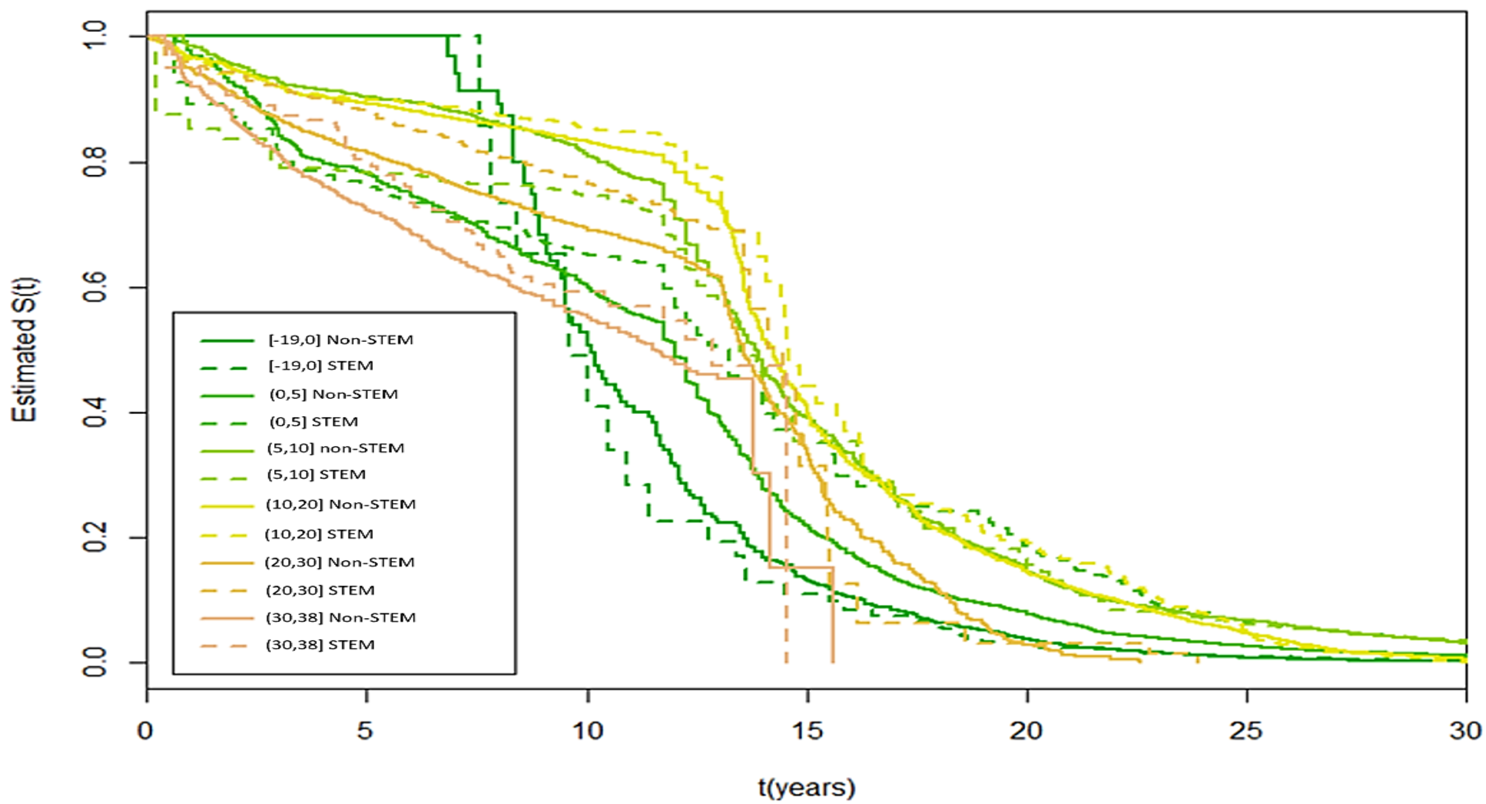

Figure 14. Estimated Survival Function by Years until Retirement Eligibility for STEM and Non-STEM Employees.

5. STEM AND PRIOR ACTIVE DUTY RESULTS

Of the entire cross-section, only 14 percent have prior active duty service. In the cross-section, there are only one percent of employees with STEM and prior active duty service. Figure 15 is the age distribution frequency by sex of the prior active duty service employees. From Figure 15, it appears that there are more young females with prior active duty service than older females with prior active duty service. The highest percentage of males with prior active duty service are in their forties, which means they could be those employees who served 20 years in active duty and then switched to federal service.

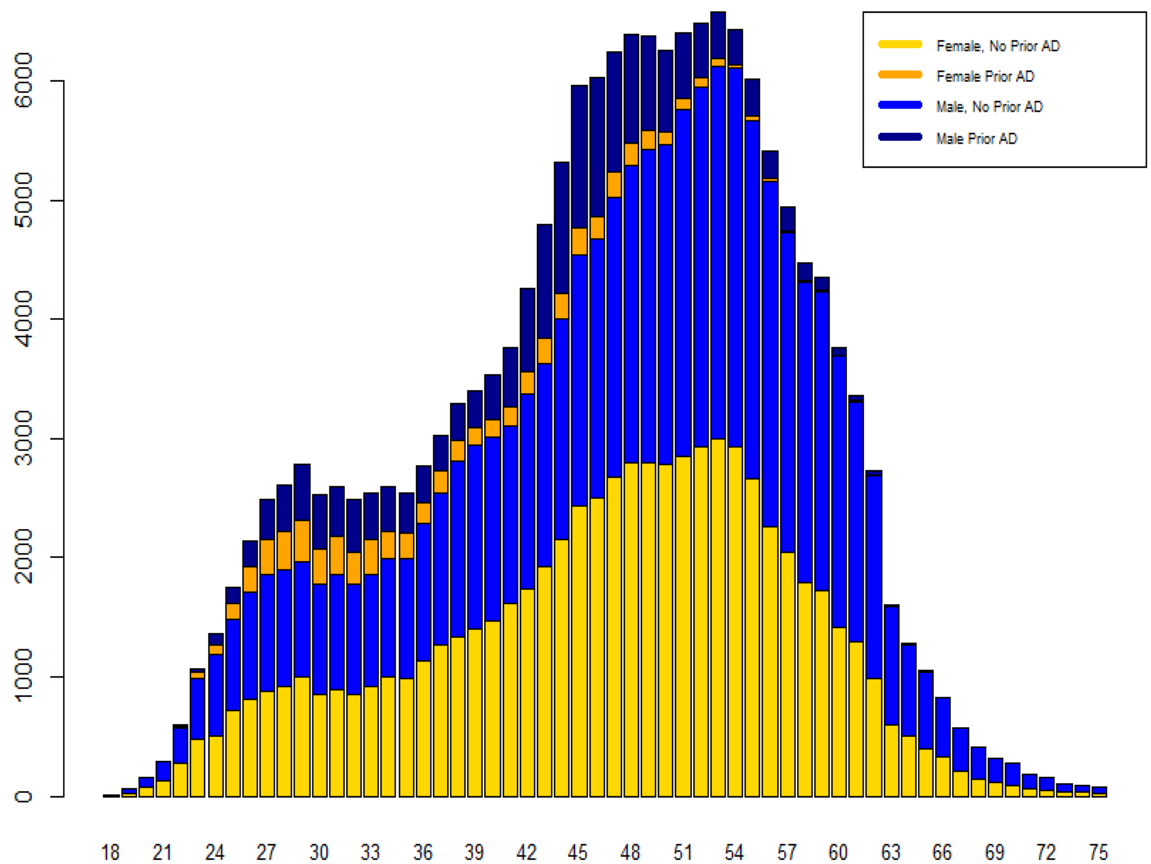

Figure 15. Age Distribution of Employees with and without Prior Active Duty Service by Sex on 2009–03–31.

Figure 16 shows the estimated survival function for those STEM employees who either do or do not have prior active duty service. From Figure 16, it appears that initially there is no difference between the estimated survival of the two until year nine. In our cross-section, it appears that there is no one who is in a STEM field with prior active duty who survives past year 15.

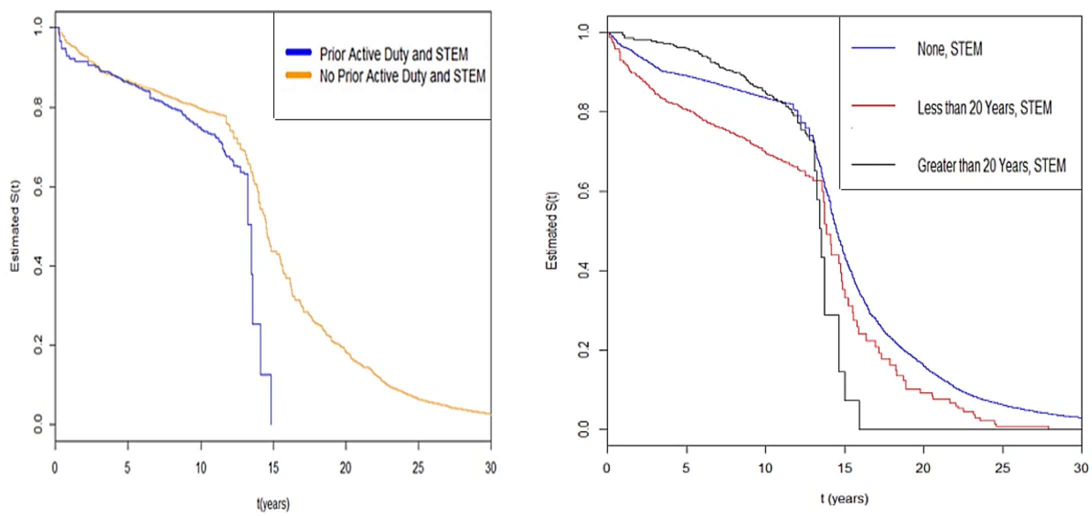

Figure 16. Estimated Survival Functions for STEM by Prior Active Duty or No Prior Active Duty Service (left) Estimated Survival Functions for STEM Employees with Either No Prior Service or Prior Service with More or Less than 20 Years (right).

Dividing the prior active duty service into greater than 20 years of active duty service and less than 20 years of service is in Figure 16. The estimated survival curves for those employees with no prior active duty service is the same for both plots. The estimated survival curve for those employees with greater than 20 years and less than 20 years of prior active duty service are different because those employees with greater than 20 years of prior active duty service survive at a higher percentage than those with less than 20 years of prior active duty service. This is true until about the 12-year mark when both estimated survival curves decrease.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. SURVIVAL TREES

In this chapter, we illustrate how survival analysis might be used to "forecast" the proportion of a cross-section expected to survive *t* more years from the cross-section date. In particular, we estimate expected proportions for the Texas and Virginia employees from our 2009 cross-section. These estimates are based on survival tree estimates of underlying survival functions for the cross-section.

Section A of this chapter gives a brief description of survival trees and our motivation for their use in this thesis. The results of the survival tree are given in Section B of this Chapter. Section C of this chapter describes a method for using the survival tree estimates for forecasting along with some of our results.

A. SURVIVAL TREES

Survival trees are a generalization of the classification and regression trees (CART) of Breiman, Freidman, Stone, and Olshen (1984). A tree partitions the data into subsets (or leaves) with a sequence of binary splits of the data. Unlike CART, a KM estimator of the survival function is computed for each leaf. We use an extension of Fu and Simonoff (2017) that accommodates both right-censored and left-truncated data.

The survival tree algorithm begins with all observations in the root node. A variable is chosen to split the root node into two subsets or nodes. Variables split the data differently depending of their type. For example, the categorical variable "Sex" splits the data according to its' levels "Male" and "Female." Whereas, a numeric variable splits the data using a splitting value. For example, the variable "Age" varies from 18 years to 75 years, so the splitting value of 20 years sends all observations with ages less than 20 into one node and all observations with ages greater than 20 to the other node. The criteria for deciding which variable to use and how to split the data using that variable is a p-value from a log-rank test of the null hypothesis that the survival distributions in the two nodes are the same. The split with the smallest p-value (the strongest evidence against the null hypothesis) is used. Splitting stops when there is no reason to continue splitting. This may be because the data in the two nodes are well modeled by a single survival function or because the data does not provide enough evidence to say they are not.

One advantage of using survival trees is that numeric variables such as “Age” do not need to be artificially partitioned in analyst-defined classes as was done in Chapter III. Optimal splits for numeric variables are chosen by the survival tree fitting algorithm. A second and the most important advantage of using survival trees is that the partition of the data is chosen objectively by the algorithm rather than subjectively, as was done in Chapter III. Finally, the third advantage to using this approach is that the survival tree gives completely non-parametric estimates of survival functions.

B. SURVIVAL TREES RESULTS

The variables used for the survival trees are “Sex,” “STEM Code,” “Education Level,” “Duty Location,” “Prior Active Duty Service,” “Years Until Immediate Retirement” and “Age” on 2009–03–31. Although “Years Until Immediate Retirement” is a numeric variable, it is treated as categorical where employees with more than 10 years until immediately retirement eligible are aggregated into a single class. For these employees, “Age” is almost perfectly linearly related to “Years Until Immediate Retirement.” As Buttrey et al. (2018) demonstrates, survival trees that treat both of these variables as numeric are more difficult to interpret because numeric “Age” and “Years Until Immediate Retirement” are interchangeable for younger employees with many years until immediately retirement eligible. Of note, “Duty Location” is not used in this tree. This indicates that having accounted for the other six variables, with this data, there does not appear to be enough evidence to see a difference in the two duty locations.

```
Surv(tstart, tstop, status) ~ Sex + StemCode.1 + EdFirst + DTY_LOC_US_ST_CD +
    yearsIRgroup + Age + priorAD

Fitted party:
[1] root
|   [2] yearsIRgroup in [-19,0], (0,1], (1,2], (2,3], (3,4], (4,5]
|   |   [3] yearsIRgroup in [-19,0], (0,1]
|   |   |   [4] Age <= 63
|   |   |   |   [5] yearsIRgroup in [-19,0]: 10.663 (n = 2423)
|   |   |   |   [6] yearsIRgroup in (0,1]
|   |   |   |   |   [7] Age <= 56: 33.002 (n = 328)
|   |   |   |   |   [8] Age > 56
|   |   |   |   |   |   [9] Age <= 58: 32.499 (n = 63)
|   |   |   |   |   |   [10] Age > 58
|   |   |   |   |   |   |   [11] Age <= 59: 23.747 (n = 298)
|   |   |   |   |   |   |   [12] Age > 59: 10.830 (n = 150)
|   |   |   [13] Age > 63: 9.499 (n = 952)
|   |   [14] yearsIRgroup in (1,2], (2,3], (3,4], (4,5]
|   |   |   [15] Age <= 60
|   |   |   |   [16] StemCode.1 in N: 12.251 (n = 3317)
|   |   |   |   [17] StemCode.1 in S
|   |   |   |   |   [18] Age <= 54: 32.091 (n = 224)
|   |   |   |   |   [19] Age > 54: 13.247 (n = 320)
|   |   |   [20] Age > 60
|   |   |   |   [21] yearsIRgroup in (1,2]: Inf (n = 406)
|   |   |   |   [22] yearsIRgroup in (2,3], (3,4], (4,5]: 6.500 (n = 157)
|   [23] yearsIRgroup in (5,10], (10,38]
|   |   [24] Age <= 31
|   |   |   [25] Age <= 26
|   |   |   |   [26] Age <= 24
|   |   |   |   |   [27] Age <= 21: 4.153 (n = 42)
|   |   |   |   |   [28] Age > 21
|   |   |   |   |   |   [29] priorAD in FALSE: Inf (n = 336)
|   |   |   |   |   |   [30] priorAD in TRUE: 8.249 (n = 57)
|   |   |   |   [31] Age > 24
|   |   |   |   |   [32] priorAD in FALSE: 14.539 (n = 339)
|   |   |   |   |   [33] priorAD in TRUE: 6.505 (n = 134)
|   |   |   [34] Age > 26
|   |   |   |   [35] Sex in F
|   |   |   |   |   [36] priorAD in FALSE: 13.301 (n = 563)
|   |   |   |   |   [37] priorAD in TRUE: 12.747 (n = 320)
|   |   |   |   [38] Sex in M
|   |   |   |   |   [39] Age <= 29: 13.659 (n = 475)
|   |   |   |   |   [40] Age > 29: 13.716 (n = 322)
|   |   [41] Age > 31
|   |   |   [42] priorAD in FALSE
|   |   |   |   [43] yearsIRgroup in (5,10]
|   |   |   |   |   [44] Age <= 51: 14.153 (n = 1003)
|   |   |   |   |   [45] Age > 51
|   |   |   |   |   |   [46] Age <= 52: 15.667 (n = 750)
|   |   |   |   |   |   [47] Age > 52
|   |   |   |   |   |   |   [48] StemCode.1 in N: 13.667 (n = 2001)
|   |   |   |   |   |   |   [49] StemCode.1 in S: 13.249 (n = 353)
|   |   |   |   [50] yearsIRgroup in (10,38]
|   |   |   |   |   [51] Age <= 40
|   |   |   |   |   |   [52] Age <= 35
|   |   |   |   |   |   |   [53] Sex in F: 13.430 (n = 516)
|   |   |   |   |   |   |   [54] Sex in M: 13.562 (n = 376)
|   |   |   |   |   |   [55] Age > 35: 14.369 (n = 1988)
|   |   |   |   |   [56] Age > 40
|   |   |   |   |   |   [57] StemCode.1 in N
|   |   |   |   |   |   |   [58] Sex in F: 13.999 (n = 3766)
|   |   |   |   |   |   |   [59] Sex in M: 14.357 (n = 2307)
|   |   |   |   |   |   [60] StemCode.1 in S
|   |   |   |   |   |   |   [61] EdFirst in 0, 1, 3: 14.083 (n = 599)
|   |   |   |   |   |   |   [62] EdFirst in 2: 16.289 (n = 523)
|   |   |   [63] priorAD in TRUE
|   |   |   |   [64] Sex in F: 13.418 (n = 1007)
|   |   |   |   [65] Sex in M
|   |   |   |   |   [66] yearsIRgroup in (5,10]: 13.086 (n = 483)
|   |   |   |   |   [67] yearsIRgroup in (10,38]
|   |   |   |   |   |   [68] Age <= 44: 15.074 (n = 1245)
|   |   |   |   |   |   [69] Age > 44: 13.304 (n = 1592)

Number of inner nodes:    34
Number of terminal nodes: 35
```

Figure 17. Estimated Survival Tree Results for Virginia and Texas Employees.

The results of the survival tree fit are summarized in Figure 17. Rather than depict the tree as a sequence of branching nodes, Figure 17 lists every tree split with a description of splitting criteria. While it is less intuitive, we find this depiction easier to extract details for this tree representation since it is a large tree with many splits.

The first split (line [2] and [23]) in Figure 17 partitions the cross-section into two nodes, those with five or fewer years until immediate retirement eligible and those with more than five years. The fact that the following several splits are based on “Age” or “Years Until Immediate Retirement” is not surprising. This confirms that both age and proximity to retirement eligibility are important predictors of separation. However, like all regression-type methods, it is difficult to ascribe causal effects to variables that are confounded.

The amount of indentation in Figure 17 indicates the depth at which the splits occur. The terminal nodes (or tree leaves) include splitting criteria, the number of observations in the terminal node, and the median number of years until separation where the median can be estimated from the estimated survival function based on the observations in that node. For example, terminal nodes [11] and [12] have estimated median separation times (from appointment date) of 23.7 years and 10.8 years, respectively. Working up the tree from nodes [11] and [12], we see that individuals in both nodes are not immediately retirement eligible on 2009–03–31, but will be within the year. The individuals in terminal node [11] are all 59 years old; whereas, the individuals in terminal mode [12] are older than 59 years old.

From the survival tree in Figure 17, we see the expected splits for “STEM Code,” “Education” and “Prior AD” but not for “Duty Location” for Virginia and Texas. The proportion of STEM employees is higher in Virginia than Texas, and the Virginia STEM employees tend to have a higher GS pay grades than the Texas STEM employees. When all of these characteristics are accounted for, the estimated survival function for a Texas employee is the same as for a Virginia employee using the survival fit of Figure 17. However, this does not mean that if the make-up of Virginia and Texas are identical, that the attrition patterns (as estimated from the estimated survival tree fit) in Virginia and Texas would be the same. Although the survival tree fit in Figure 17 may impart some insight, we use it in this chapter primarily as a tool for prediction.

C. FORECASTING AND CROSS VALIDATION

The survival functions as displayed in Chapter II and those from the survival trees estimate the probability of surviving at least *t* years from appointment date, but they do not give the probability that an individual in the 2009–03–31 cross-section will survive an additional *t* years. The probability of surviving *t* additional years is a conditional probability. For an individual who has been employed by DA for s years on 2009–03–31 (as indicated by "Start Age"), the conditional probability of surviving an additional *t* years is given by the equation $S(s+t)/S(s)$.

Given the individuals' "Age," "Sex," etc., the survival tree fit of Figure 17 estimates the individuals' survival function which can then be used to estimate the probability that the individual survives additional *t* years past his "Start Age," s. The cross-section, though, is composed of a mixture of individuals with different "Start Ages" and different estimated survival functions. To estimate the proportion of individuals from the cross-section (or from a subset of the cross-section) who will survive *t* years from 2009–03–31 requires estimating the probability of surviving at least *t* more years for each individual in the cross-section. The estimated expected proportion of the cross-section who survive *t* additional years is then the sum of the individuals' estimated probabilities.

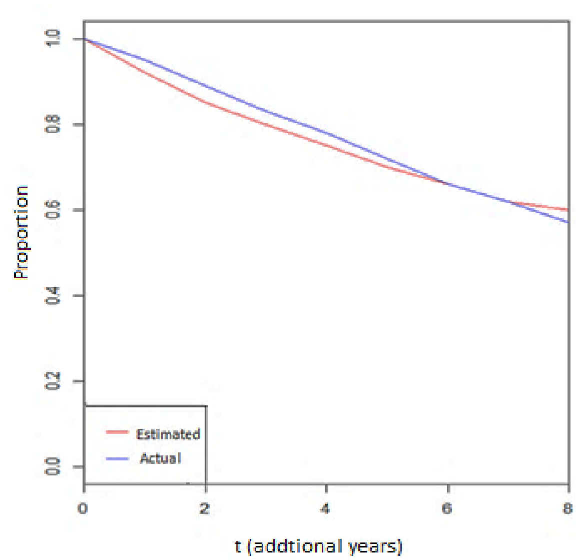

Figure 18. Estimated Proportion and Actual Proportion as of 2009–03–31 of the Cross-Section Who Survive an Additional *t* Years.

Figure 18 plots the estimated proportion of the entire cross-section who survive an additional *t* years from 2009–03–31 in blue. As a check, we also plot (Figure 18) the actual proportion who survive at least *t* additional years in red. The estimated proportion is quite similar to the actual proportion. This is comforting, but it is not a true test of the ability of our models to forecast because the same set of data is used to construct the survival tree and compute the actual proportions.

A better approach for assessing the reliability of the forecasts is to validate the forecasts on an independent set of data. To do this, we use cross-validation in the next section.

1. CROSS-VALIDATION

We use 10-fold cross-validation to see how the method described in the previous section might perform on an independent set of data from a similar time period to check the validity of our model. In the 10-fold cross-validation, about 10 percent of the 29,735 observations are randomly selected for each fold. For each fold we fit a survival tree to the remaining 9 folds, yielding a total of 10 survival trees. Then we test each of the 10 survival tree fits by estimating the proportion of individuals from the excluded fold who survive an additional *t* years.

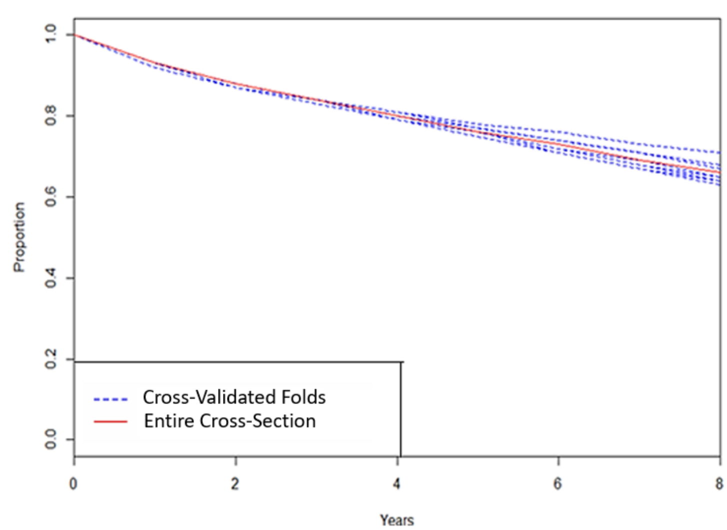

Figure 19. Cross-Validation of the Estimated Cross-Section Proportion who Survive an Additional *t* Years Past 2009–03–31.

The results of these 10 estimated proportions are the blue dashed lines in Figure 19. The red line in Figure 19 are the estimated proportions for the entire data. The results of the cross-validation indicate uncertainty in the estimated proportion who survive an additional *t* years. There appears to be less variability the first three years than after the three-year mark; there is increasing variability until the eight-year mark. The overall estimated probability decreases similar to the estimated proportion calculated from Figure 18.

2. STEM AND NON-STEM IN VIRGINIA AND TEXAS

In this section, we forecast the proportion of individuals who survive up to *t* years for *t* less than eight years and for several of the subsets investigated in Chapter III. The survival functions displayed in the previous chapter and those from the survival tree estimate the probability of surviving at least *t* years from appointment date.

Figure 20 forecasts the proportion who survive *t* more years of STEM and non-STEM employees by their duty location of either Texas or Virginia. When we forecast eight years, there does not appear to be a difference for Texas and Virginia for both the STEM and non-STEM employees.

The forecasting results from Figure 21 are similar to those in Figure 8. The STEM males and females' survival probability is similar, and for the non-STEM employees, the males survive at slightly higher percentages than the females.

Consistent with the estimated survival function in Figure14, those employees who are immediately retirement eligible are forecasted in Figure 22 survive the worst out of all of the groups that are immediately retirement eligible.

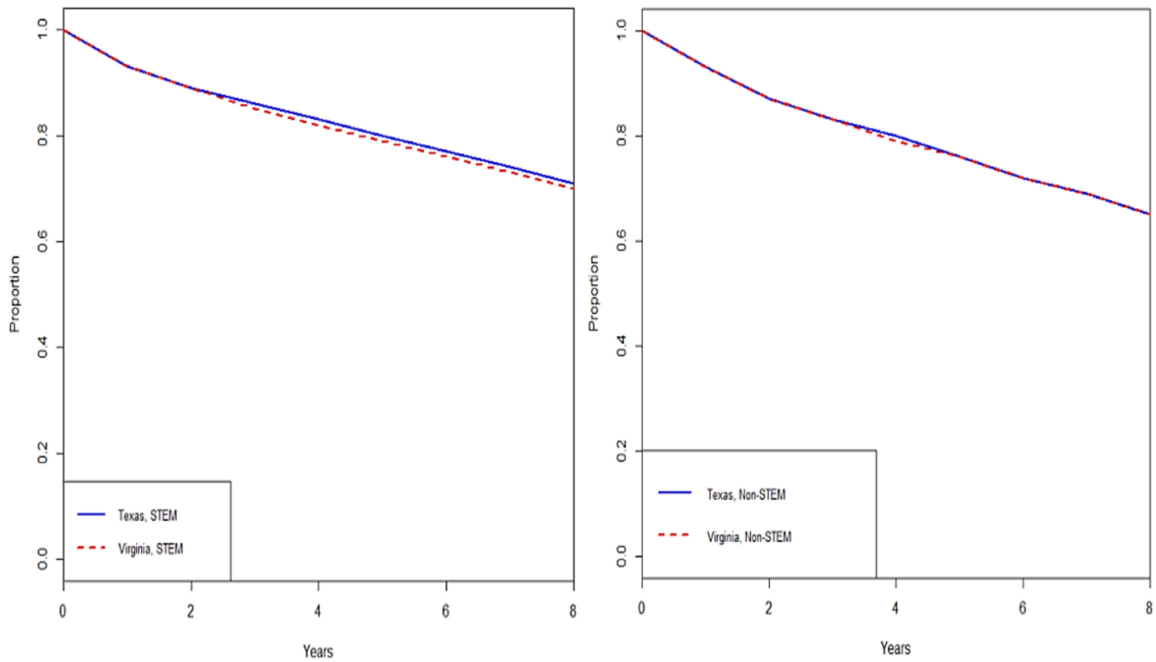

Figure 20. Forecasting the Proportion of the 2009 Cross-Section Who Survive t years beyond 2009–03–31 by Duty Location and STEM (left) and Non-STEM (right).

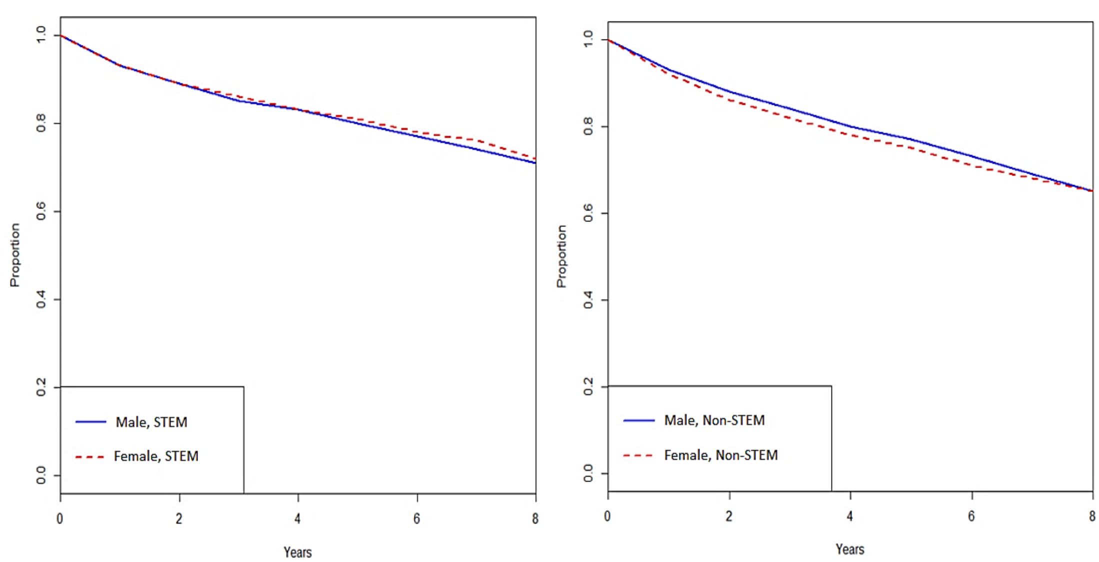

Figure 21. Forecasting the Proportion of the 2009 Cross-Section who Survive t Years beyond 2009–03–31by Sex and STEM (left) and Non-STEM (right).

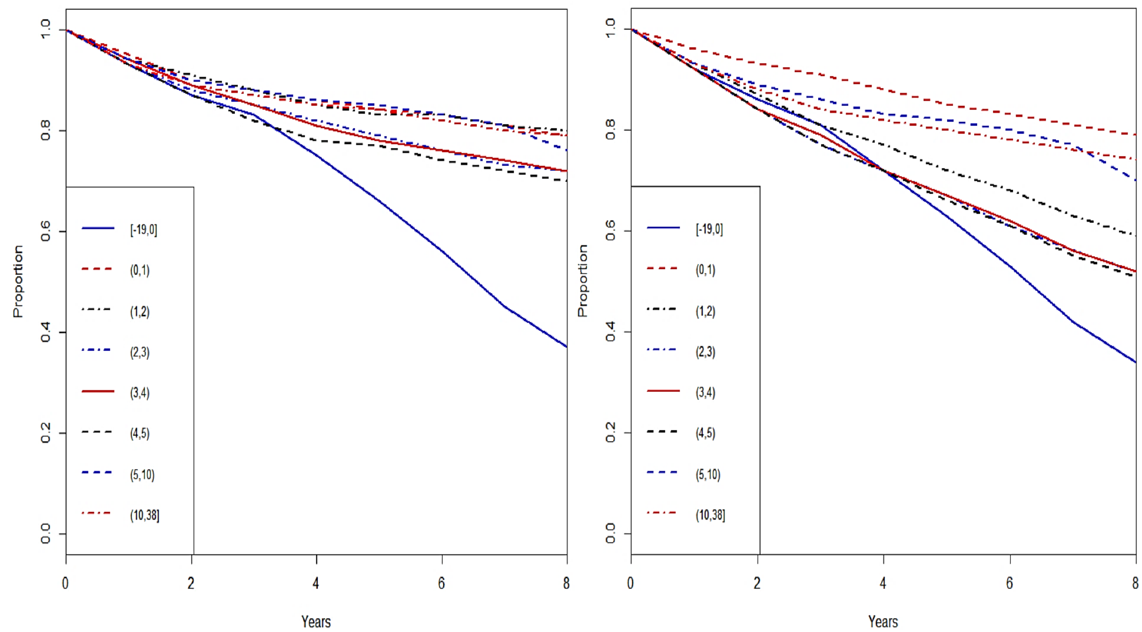

Figure 22. Forecasting the Proportion of the 2009 Cross-Section Who Survive *t* Years beyond 2009–03–31 by STEM (left) and Non-STEM (right).

For both STEM and non-STEM in Figure 22, the employees who have 10–38 years until immediate retirement eligibility survive relatively better than some of the other groups. Surprisingly, for non-STEM employees, in Figure 22, the employees with highest percentage of survivability after eight years are those with zero or one year until retirement eligible.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND FUTURE WORK

In this thesis, we model attrition of full-time, permanently appointed, GS DA civilian employees with factors of interest from our sponsor using KM estimated survival functions. We do this in order to understand attrition factors related to the STEM DA workforce and how these might differ for DA employees with non-STEM occupations. We also build models to forecast attrition with these variables through the use of survival trees.

A. FINDINGS

In Chapter III, we estimate survival functions based on subsets of a cross-section of civilians employed by DA on 2009–03-31 to see the estimated survival probability as a function of time from appointment. We learn that the attrition of these employees appears to be almost piecewise linear with decreasing attrition rate until about the 12-year mark, after which, there is a drastic increase in attrition. We are unable how much of this drop is due to an increase in attrition after 12 years and how much is due to how we approximate appointment date for DA employees appointed prior to 2005. However, with time as civilian DoD transaction files become available, we will be able to confirm the results.

We also learn that there is no difference in attrition between males and females in STEM fields, but there is a difference in attrition between males and females in non-STEM fields. This finding is consistent with the study of new hires from Buttrey et al. (2018). Despite a lack of differences of attrition of the males and females in STEM positions, there does appear to be a difference in attrition of the STEM employees by geographic location in Texas and Virginia. Notably, the Virginia STEM employees have higher attrition rates than the Texas STEM employees. The high attrition rates of Virginia STEM employees could be due to STEM job availability in Virginia, which was expected to grow eight percent between 2008 and 2018, according to a 2014 Georgetown University study by Carnevale, Smith, and Melton. The STEM jobs increase in Virginia is significantly higher than the increase Texas saw during the same time period. This implies a potentially greater demand for STEM capabilities within the Virginia and national capital region, which may be beneficial to highly qualified STEM employees that could see multiple organizations competing to recruit them.

Another explanation for higher attrition of STEM employees in Virginia than in Texas could be due to the location of the STEM jobs in each state. As seen in Figure 10, the density of the STEM jobs in Virginia is much higher than the density of the Texas STEM jobs. Also note that while the comparisons in this thesis only consider Virginia, Figure 10's identification of STEM hubs shows the concentration of jobs in Virginia spreads into D.C. and other neighboring states. While Texas may have numerous STEM job opportunities, the cost to the employee of transferring from a STEM position in Dallas to Houston is substantially more than transferring within the state of Virginia or to the neighboring D.C. area. The latter may entail a change in commute, but with no need to uproot a household for a more disruptive move across the state. Therefore, the higher density of STEM opportunities may create more job mobility for highly qualified employees within the Virginia, D.C. area.

We also find that the difference in attrition of the STEM employees in Texas and Virginia are the most different in the first three years, after which their attrition rates look more similar. We do not know exactly why these employees are leaving the DoD STEM positions in Virginia within the first three years, but we hypothesize it could be due to numerous, easily transitioned to alternative opportunities enticing employees to pursue something new. Additionally, given the concentration of STEM jobs, employees may seek positions within the Virginia area to establish their career at an entry-level position prior to departing for other opportunities after two to three years. This behavior could be addressed through policy implementation, such as locality pay increases to incentivize retention, or incurring retention obligations associated with certain positions, training, bonuses, or allowances.

Though there is a difference in attrition of STEM employees by geographic location, there does not appear to be much difference in attrition by geographic location for non-STEM employees. This could be due to the small increase of non-STEM jobs of 4.5 percent from 2005 to 2015 versus STEM jobs, which increased 9.8 percent over this timeframe (Simmons 2016).

In Chapter IV, the results indicate that age and time until immediately retirement eligible are the most important factors for predicting attrition. This finding is consistent with the life cycle stability hypothesis from the study of Moynihan and Pandey (2007). Their hypothesis suggests that older employees who have been in a position for a long time are negatively correlated with turnover (Moynihan and Pandey 2007). Thus, despite the differences seen in Chapter III survival analyses results about differences in attrition by duty location, age and time until immediately retirement are better predictors of attrition. This also implies further policy considerations for organizational leadership; specifically, the nuances of retirement policies and benefits shape the workforce and its retention behavior over time. Organizational leaders and policymakers may benefit from discussions on whether the observed retention behavior is beneficial or not to the organization, as well as how to reconfigure retirement and retention incentives accordingly.

B. RECOMMENDATIONS FOR FUTURE WORK

During the course of our work, several follow-on research opportunities become apparent. The DMDC data stored in the PDE contains other variables that should be considered. The most immediately important are ZIP codes of the employees and their UIC. Using these variables combined with local demographic and economic information may enable an understanding of attrition for specific geographic locations and specific units while controlling for the local economy, as most of the other turnover literature does. The controls in this thesis perform well with the employee life cycle, but do not get at external competing demand for federal employment and employee satisfaction.

Additionally, analysis of the civilian billet file, giving numbers and types of jobs for each DoD agency, available through the PDE, could provide important recommendations for a unit to understand the attrition of their employees by specific jobs. The work with the billet file could help with understanding manpower planning within the organization. No one has yet to analyze the billet file in conjunction with civilian personnel records, so any findings would be new and would benefit the DoD.

Follow-on work should also include a comparative analysis of this cross-section with cross-sections from other years to understand the estimated survival functions at different times. This is particularly important because 2009 is a recession-era year. These studies will catalogue changes in attrition patterns with time and might help DoD better understand the influences of policy changes and economic factors.

A final recommendation for follow-on work is a comparative analysis of different geographic locations. Texas and Virginia are chosen for this thesis, in part, because they were two of the states with the highest number of STEM employees in our cross-section. Other states with high STEM DA employment, such as California, should be studied to compare say trends of the private and public sector STEM jobs. Furthermore, by using ZIP codes to identify comparable regions (such as metropolitan or rural areas) rather than relying strictly on state boundaries, other interesting comparisons of attrition can be made.

APPENDIX A. STEM CODES FOR STEM, MEDICAL, AND SOCIAL SCIENCE

The STEM codes are classified by OPM in to the following groups (OPM 2018):

- “Regular” stem jobs have codes “0401,” “0403,” “0405,” “0408,” “0410,” “0413,” “0414,” “0415,” “0430,” “0434,” “0435,” “0437,” “0440,” “0454,” “0457,” “0460,” “0470,” “0471,” “0480,” “0482,” “0485,” “0486,” “0487,” “0801,” “0803,” “0804,” “0806,” “0807,” “0808,” “0810,” “0819,” “0828,” “0830,” “0840,” “0850,” “0854,” “0855,” “0858,” “0861,” “0871,” “0880,” “0881,” “0890,” “0893,” “0896,” “1301,” “1306,” “1310,” “1313,” “1315,” “1320,” “1321,” “1330,” “1340,” “1350,” “1360,” “1370,” “1372,” “1373,” “1382,” “1384,” “1386,” “1501,” “1510,” “1515,” “1520,” “1529,” “1530,” “1541,” “1550,” or “2210.”
- “Medical” jobs have codes “0601,” “0602,” “0610,” “0620,” “0630,” “0631,” “0633,” “0635,” “0637,” “0638,” “0639,” “0651,” “0660,” “0662,” “0665,” “0667,” “0668,” “0670,” “0682,” “0685,” “0690,” and “0696.”
- “Social Science” jobs have codes “0101,” “0150,” “0180,” “0184,” “0190,” or “0193”

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. DISTRIBUTIONS OF STEM AND NON-STEM 2009 TEXAS AND VIRGINIA CROSS-SECTION OF EMPLOYEES BY VARIABLE

		VA (STEM)	TX (STEM)	VA (Non-STEM)	TX (Non-STEM)
GS PAY GRADE	GS02	NA	NA	1 (0.008)	31 (0.25)
	GS03	NA	NA	15 (0.11)	41 (0.34)
	GS04	NA	NA	139 (1.05)	987 (8.08)
	GS05	4 (0.16)	4 (0.23)	446 (3.37)	1219 (9.97)
	GS06	NA	NA	528 (3.99)	990 (8.10)
	GS07	36 (1.40)	36 (2.11)	1151 (8.70)	1347 (11.02)
	GS08	NA	NA	224 (1.69)	284 (2.32)
	GS09	92 (3.57)	149 (8.74)	992 (7.50)	1307 (10.69)
	GS10	NA	NA	35 (0.26)	42 (0.34)
	GS11	157 (6.09)	188 (11.03)	900 (6.80)	1184 (9.69)
	GS12	425 (16.47)	284 (16.67)	1011 (7.64)	701 (5.74)
	GS13	204 (7.9)	75 (4.40)	666 (5.03)	173 (1.42)
	BAND	1662 (64.41)	968 (56.81)	7122 (53.82)	3915 (32.04)
SEX	MALE	1798 (69.69)	1313 (77.05)	6324 (47.80)	6341 (48.11)
	FEMALE	782 (30.31)	391 (22.95)	6906 (52.20)	5880 (51.89)
AGE	[19,29]	191 (7.40)	134 (7.86)	1323 (10.00)	1273 (10.42)
	(29,34]	129 (5.00)	113 (6.63)	1478 (11.17)	1414 (11.57)
	(34,39]	189 (7.32)	152 (8.92)	1629 (12.31)	1197 (9.79)
	(39,44]	291 (11.28)	192 (11.27)	1032 (7.80)	1290 (10.56)
	(44,47]	280 (10.85)	179 (10.50)	1663 (12.57)	1349 (11.04)
	(47,49]	207 (8.02)	128 (7.51)	1148 (8.68)	913 (7.47)
	(49,52]	349 (13.53)	194 (11.38)	1155 (8.73)	1442 (11.80)
	(52,55]	323 (12.52)	200 (11.74)	1429 (10.80)	1238 (10.13)
	(55,59]	323 (12.52)	205 (12.03)	1157 (8.74)	999 (8.17)
	(59,75]	298 (11.55)	207 (12.15)	1216 (9.19)	1106 (9.05)
FIRST FILE DATE	≤ 2005	2956 (84.94)	1320 (77.46)	10317 (77.98)	9072 (74.26)
	2006	111 (3.19)	87 (5.11)	658 (4.97)	766 (6.30)
	2007	122 (3.51)	70 (4.11)	767 (5.80)	850 (6.96)
	2008	231 (6.64)	176 (10.33)	1188 (8.98)	1269 (10.39)
	2009	60 (1.72)	51 (2.99)	300 (2.27)	260 (2.13)
PRIOR ACTIVE DUTY SERVICE	True	242 (7.45)	185 (7.79)	2434 (18.40)	2528 (20.69)
	False	3008 (92.55)	2189 (92.21)	10796 (81.60)	9693 (79.31)
HIGHEST RECORDED	Less than High School	8 (0.31)	1 (0.06)	84 (0.63)	107 (0.88)

EDUCATION LEVEL	High School Diploma	877 (33.99)	492 (28.87)	7654 (57.85)	8902 (72.84)
	College Degree	1044 (40.47)	859 (58.41)	3084 (23.31)	2035 (16.65)
	Master's Degree	651 (25.23)	352 (20.66)	2408 (18.20)	1177 (9.63)
YEARS UNTIL IMMEDIATE RETIREMENT ELIGIBLE	[-19,0]	325 (12.60)	201 (11.80)	1501 (11.35)	1346 (11.01)
	(0,1]	66 (2.56)	51 (2.99)	408 (3.08)	316 (2.59)
	(1,2]	123 (4.77)	73 (4.28)	642 (4.85)	515 (4.21)
	(2,3]	86 (3.33)	51 (2.99)	473 (3.58)	395 (3.23)
	(3,4]	67 (2.60)	67 (3.93)	451 (3.41)	379 (3.10)
	(4,5]	109 (4.22)	58 (3.40)	513 (3.88)	422 (3.45)
	(5,10]	416 (16.12)	250 (14.67)	2165 (16.36)	1822 (14.91)
	(10,38]	1388 (53.80)	953 (55.93)	7077 (53.49)	7026 (57.49)

LIST OF REFERENCES

Asch B (2002) Paying, promoting and retaining high-quality workers in the Department of Defense Research brief. RAND Corporation, Santa Monica. https://www.rand.org/pubs/research_briefs/RB7550.html.

Breiman L, Freidman J, Stone C, Olshen R A (1984) *Classification and Regression Trees*, (Wadsworth International Group, Belmont, CA).

Buttrey S, Klingensmith K, Whitaker L (2018) Civilian Department of Defense attrition. Technical report TACC-M-TR-18-012, U.S. Army TRADOC Analysis Center, Monterey, CA.

Butz W, Kelly T, Adamson D, Bloom G, Fossum D, Gross M (2004) Is the federal scientific and technical workforce facing a crisis? RAND Corporation, Santa Monica. https://www.rand.org/pubs/research_briefs/RB1505.html.

Carnevale A, Smith N, Melton M (2014) STEM state-level analysis. Accessed March 1, 2019, https://cew.georgetown.edu/wp-content/uploads/2014/11/stem-states-complete-update2.pdf.

Copeland C (2011) The federal workforce: characteristics and trends. CRS report RL34685, Congressional Research Service, Washington, DC. https://www.everycrsreport.com/reports/RL34685.html.

Fu W, Simonoff J (2017) Survival trees for left-truncated and right-censored data, with application to time-varying covariate data. *Biostatistics* 18(2): 352–369, https://doi.org/10.1093/biostatistics/kxw047.

Jeffrey T (2014) Female share of federal workforce declining under Obama. CNS NEWS, https://www.cnsnews.com/commentary/terence-p-jeffrey/female-share-federal-workforce-declining-under-obama.

Kaplan E, Meier P (1958) Nonparametric estimation from incomplete observations. *J. of the Am. Statis. Assoc.* 53(282): 457–481, https://doi:10.2307/2281868.

Kleinbaum D, Klein M (2005) *Survival Analysis: A Self-learning Text*, 2nd ed. (SpringerLink Books, Springer, NY).

Leider J, Harper E, Shon W, Sellers K, Castrucci B (2016) Job satisfaction and expected turnover among federal, state, and local public health practitioners. *Am. J. of Pub. Hlth.* 106(10): 1782–1788, doi:10.2015/AJPH.2016.303305.

Moynihan D, Pandey S (2007) The ties that bind: Social networks, person-organization value fit, and turnover intent. *J. of Pub. Adm. Res. and Thry.* 18(2): 205–227, doi:10.1093/jopart/mum013.

Nataraj S, Hanser L, Camm F, Yeats J (2014) The future of the Army's civilian workforce: comparing projected inventory with anticipated requirements and estimating cost under different personnel policies. RAND Corporation, https://www.rand.org/pubs/research_reports/RR576.html.

National Academy of Engineering and National Research Council (2012) *Assuring the U.S. Department of Defense a Strong Science, Technology, Engineering and Mathematics (STEM) workforce*. (The National Academies Press, Washington, DC).

Office of Personnel Management (2018) General Schedule Classification and Pay. Accessed December 2, 2018, https://www.opm.gov/policy-data-oversight/pay-leave/pay-systems/general-schedule/.

Office of Personnel Management (2019) Handbook of Occupational Groups and Families. Accessed January 10, 2019, https://www.opm.gov/policy-data-oversight/classification-qualifications/classifying-general-schedule-positions/occupationalhandbook.pdf.

Rothwell J (2013) The hidden STEM economy. Accessed February 20, 2019, https://www.brookings.edu/research/the-hidden-stem-economy/.

Simmons S (2016) STEM job growth expected to exceed growth in non-STEM jobs over the next decade. Accessed March 5, 2019, https://chmuraecon.com/blog/2016/february/29/stem-growth-expected-to-exceed-growth-in-non-stem-jobs-over-the-next-decade/.

Vocational Rehabilitation and Employment Service Manual (2014) Veterans' preference and special hiring authorities. Accessed March 1, 2019, https://benefits.va.gov/WARMS/docs/admin28/M28R/Part_VI/pt06_ch08_secA.pdf.

Wadsworth L, Llorens J, Facer R (2018) Do workplace flexibilities influence employment stability? An analysis of alternative work schedules, turnover intent and gender in local government. *Intl. J. of Org. Thry. & Behav.* 21(4): 258–274, https://doi.org/10.1108/IJOTB-07-2018-0077.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California